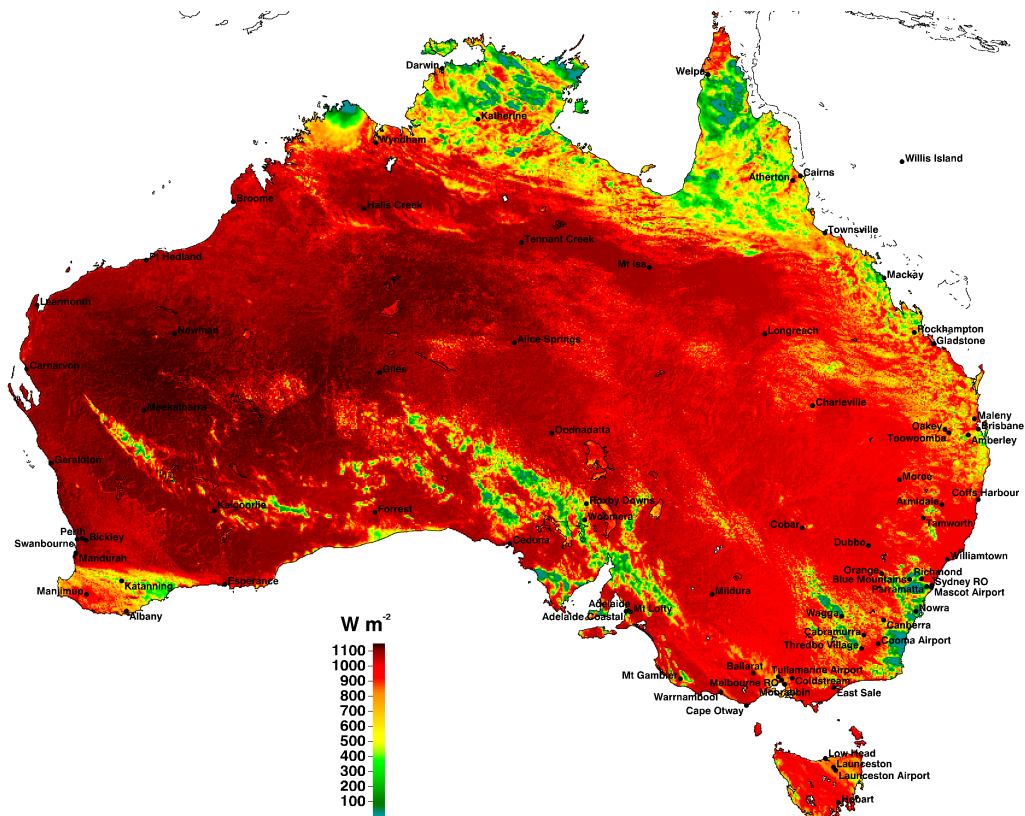


# NatHERS 2019 Climate File Update

*Prepared for NatHERS Administrator*

*May 2020*



Global irradiance at (nominal) 0300 UT, 1 January 1998, as derived from satellite data by the Bureau of Meteorology, overlaid with NatHERS reference site locations.

Prepared by:  
Ben Liley




For any information regarding this report please contact:

Ben Liley  
Atmospheric Scientist  
Atmospheric Processes  
+64-3-440 0427  
Ben.Liley@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd  
Lauder  
Private Bag 50061  
Omakau 9352

Phone +64 3 440 0055

NIWA CLIENT REPORT No: 2019175WN  
Report date: May 2020  
NIWA Project: DCE19601

Quality Assurance Statement		
	Reviewed by:	David F. Pollard
	Formatting checked by:	Alex Quigley
	Approved for release by:	Steve Wilcox

# Contents

<b>Executive summary .....</b>	<b>5</b>
<b>1 Background .....</b>	<b>6</b>
1.1 History .....	6
1.2 Comparison of 2005 and 2016 datasets .....	7
1.3 Climate file update and review with Bureau of Meteorology .....	7
1.4 EnergyPlus TMY files .....	7
<b>2 Climate station data .....</b>	<b>8</b>
2.1 Primary locations.....	8
2.2 Additional sites .....	17
2.3 Cloud data .....	17
2.4 Wind data .....	18
<b>3 Data quality control.....</b>	<b>19</b>
3.1 Error detection .....	19
3.2 Missing data .....	19
3.3 Visible anomalies.....	19
3.4 Statistical anomalies.....	19
3.5 Tropical cyclones .....	22
3.6 Drying of wet bulb wick.....	23
3.7 Wind data .....	24
<b>4 Solar radiation.....</b>	<b>25</b>
4.1 Ground-based data .....	25
4.2 Errors in ground-based radiation data .....	27
4.3 Satellite-derived radiation data.....	28
4.4 Spatial interpolation of erroneous data .....	29
4.5 Comparison with ground-based data .....	30
4.6 Himawari 8 data .....	32
<b>5 Derivation of RMY/TMY files .....</b>	<b>34</b>
5.1 Finkelstein-Schafer statistics .....	34
5.2 Ambiguity in the Sandia method .....	35
5.3 EnergyPlus TMYs .....	35

6	Review with Bureau of Meteorology .....	36
7	References .....	38
Appendix A	Example Time Series .....	39
Appendix B	Finkelstein-Schafer Statistic.....	52
Appendix C	Monthly Mean and S.D. of $G, T, T_d, W_s$ .....	55



## Executive summary

The Australian Nationwide House Energy Rating Scheme (NatHERS) uses computer software to simulate household energy efficiency for regulatory purposes. The software makes use of climate files in a prescribed format to represent the many climates across the country. The present system uses 69 climate zones, with Representative Meteorological Year (RMY) data files developed in 2005 using data from 1969 to 2004 inclusive.

Subsequent revisions increased the number of zones to 83, and used data up to 2007, 2011, and 2015 respectively to create new RMYs. The latter revisions have had the benefit of solar radiation data derived at high resolution from satellite data. The 2016 revision also discontinued use of data before 1990, for reasons of data quality and to better reflect any recent changes in climate.

In 2019, NIWA was contracted to further study the 2016 files and update them to the end of 2018.

Since 2016, reference simulations by CSIRO showed that impacts of changing the climate files in NatHERS varied across Australia, with effects on regulation. The NatHERS Steering Committee decided not to adopt any revisions until the National Construction Code (NCC) 2022 update cycle, and use the intervening time to understand the changes, confirm their validity, and adjust rating systems accordingly. Within the new contract, we analysed changes in Melbourne, Perth, and Brisbane data, and produced a pair of RMY files separately based on ground-based and satellite estimates for Tullamarine. That work was reported in May 2019.

This report describes the subsequent update, NatHERS 2019, using new data to the end of 2018 from the Bureau of Meteorology. We describe the quality control, and the creation of files of the same data in EnergyPlus format, widely used by software for simulating building energy performance.

The 2019 update did not add any new climate zones, nor examine or redefine any boundaries. The reference sites were unchanged, and no new synoptic (three-hourly or less frequent, staffed) sites were needed, but four new automatic weather stations (AWSs) were used. Three of those were Bureau of Meteorology upgrades of the AWS used previously.

Quality control on all data is detailed and extensive, and it has been expanded and refined over a decade. The full suite of error checks is described, together with a newly-discovered issue with a small fraction of the ground-based solar radiation data.

The gridded solar radiation data used for most sites, and to complete time series for radiation sites, are derived from geostationary satellite imagery. Since 22 March 2016, the images have come from the Himawari 8 satellite. When we added those data to the time series for the NatHERS sites, we found them to be inconsistent with earlier data at some sites, especially those in northern Australia.

The 2016 files were recreated in EnergyPlus format, including a required shift in time alignment for the radiation data, and the addition of rainfall codes.

In November 2019, Ben Liley visited the Bureau of Meteorology in Melbourne to review the data choices, error detection, and processing of Bureau data into time series for NatHERS. This was to allow the possibility that the Bureau might implement the procedures to directly supply data to the same level of quality control. The record of corrections to ground-based radiation data was gratefully received, as was the analysis of the change in gridded data with Himawari 8 by Dr Ian Grant, who was responsible for the satellite image processing into radiation data over the last two decades. Very sadly, Ian passed away on 30 November 2019, after a nine-month battle with cancer.

# 1 Background

## 1.1 History

Australia's Nationwide House Energy Rating Scheme (NatHERS) uses computer software to simulate household energy efficiency for regulatory purposes. At the core of NatHERS software is the 'Chenath Engine', developed by CSIRO (e.g., Chen 2016). To encompass the range of climate types and consequent energy demands across Australia, NatHERS represents the country with climate zones, distributed nationally and 69 in number at present. Each zone has an associated climate data set, originally from the Australian Climatic Data Bank (ACDB), being hourly time series of:

- air temperature
- moisture content
- pressure
- wind speed
- wind direction
- cloud cover
- global (horizontal) solar irradiance
- diffuse irradiance
- direct radiation (on a sun-tracking surface).

Zones are attributed to a single location, usually of a Bureau of Meteorology Automatic Weather Station (AWS), but for a number of zones additional sites are needed to complete the range of data types or extend the coverage. Because the time series can be of different duration and have missing or erroneous data, the actual representation of each site is by a reference meteorological year (RMY) produced for each zone from a statistically-determined subset.

Derivation of NatHERS RMYs is based on the method for Typical Meteorological Years (version 2; TMY2) described by Marion and Urban (Marion and Urban 1995) with the weightings suggested therein. The RMY designation, rather than TMY, denotes that NatHERS data files are presented in (an amended version of) the fixed record format of ACDB, as described by Delsante (2005). In contrast, TMY files are generally promulgated in the comma-separated-variable format of EnergyPlus Weather (\*.EPW) files, with different representation of humidity and potentially many additional data fields.

The climate files currently used in NatHERS were implemented in 2005 and comprise climate data from 1969 to the end of 2004. The time series and specified 69 zones were derived from an update of the ACDB, originally developed for 18 climate zones (Walsh, Munro et al. 1983). In a series of updates, the number of zones was expanded to the present 69, and recent versions now cover 83 zones.

The latest (2016) revision (Liley, J Ben 2017) departed from previous work by using Bureau of Meteorology data just since 1990, coincident with satellite-derived solar radiation data available for the continent. Mostly, the ground-based data are from AWS sites that report half-hourly, but three sites draw primarily on data from 'synoptic' (staffed; three-hourly, six-hourly, or twice-daily) data, as

previously. New data sources were added where the record for previous sites was inconsistent or discontinued. The Willis Island dataset was not updated, as it is not covered in the satellite-derived solar radiation data. For future updates, a new station was created for Christmas Island (#83 XI), based on data for the nearby Cocos Island which, as a station in the international Baseline Surface Radiation Network (BSRN), has solar radiation data of the highest quality.

Similar high quality ground-based data are available for 18 other NatHERS sites, but for some of them the record is short. Those data were used in the time series when available, and the satellite data for those sites were compared with the ground-based data for the periods in common.

## 1.2 Comparison of 2005 and 2016 datasets

Analysis by NatHERS of how the proposed adoption of the 2016 revision would affect energy ratings for specified house designs showed some differences from use of the 2005 datasets. We analysed differences between the two datasets in year selection and how that affected the representation of Climate Zones #21 ME: Melbourne RO, #13 PE: Perth, and subsequently #10 BR: Brisbane (Liley, Ben 2019). To enable an assessment of validity of the satellite-derived radiation data, we produced two versions of the 2016 RMY for #60 TU: Tullamarine, for which there were sufficient ground-based data. As summarised in Liley (2019), testing with these data files by CSIRO showed minimal difference in heating and cooling loads, separately and combined, for the benchmark house design.

## 1.3 Climate file update and review with Bureau of Meteorology

The principal requirement of this work was the update of NatHERS climate files from the 2016 version, based on 1990 – 2015, to include data up to the end of 2018. A further contract requirement was that Ben Liley travel to Melbourne to review with Bureau of Meteorology staff the data choices, quality control algorithms, and further analyses, sufficient to enable the possibility of the Bureau in future directly supplying data in a form ready for creating RMYs and TMYs.

## 1.4 EnergyPlus TMY files

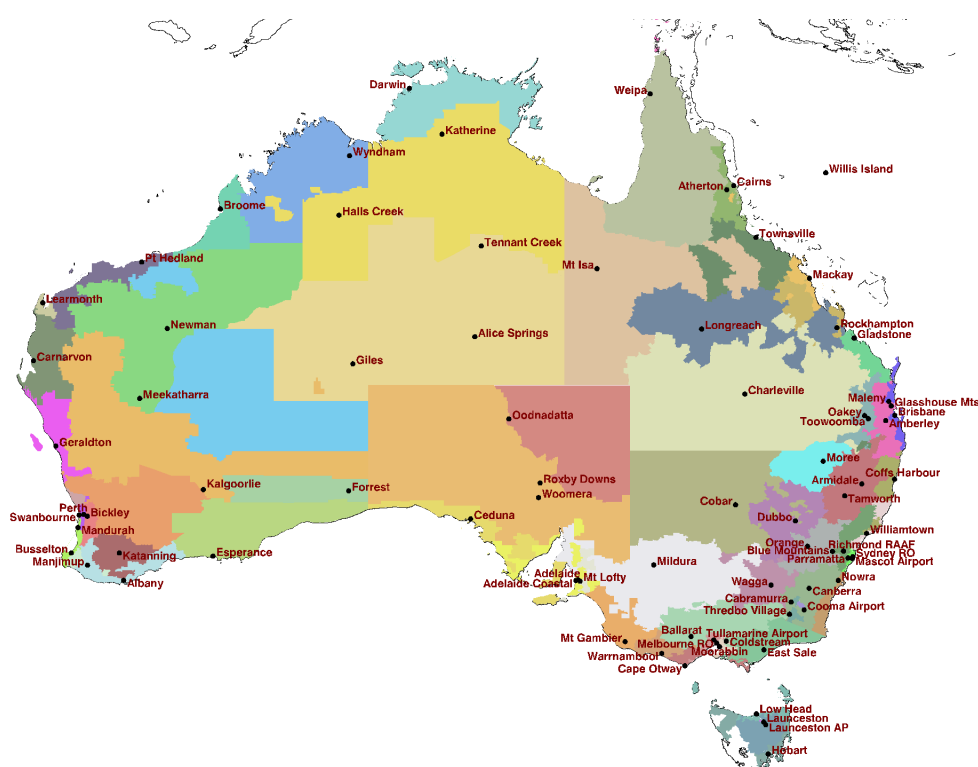
An additional requirement was for conversion of NatHERS climate files to the EnergyPlus format, widely used by researchers in building energy simulation. The NatHERS 2012 data files in EnergyPlus ('TMY') format have over 250 registered users. As described below, the NatHERS 2016 files, rather than the subsequent 2019 set, are expected to be served to users, though both are available.

## 2 Climate station data

### 2.1 Primary locations

There are 69 zones in the present version of NatHERS, based on those in the revision of ACDB to 2004. The zones, as of the latest (2013) revision, and their boundaries in terms of post codes, are shown in the coloured polygons in Figure 1. They cover the Australian continent, with higher density in populous areas. Also marked and labelled in Figure 1 are 82 of the 83 sites (excluding #83 XI: Christmas Island) for which climate time series and RMYs have been developed.

All sites have meteorological instruments to record many of the climate parameters used in building energy simulation. In particular, dry and wet bulb temperatures are recorded at all stations, or at adjacent sites, sufficient to provide complete records after temporal or spatial interpolation. Most sites record atmospheric pressure, though it is relatively unimportant. Air pressure is required for conversion between the different representations of atmospheric water vapour; wet bulb temperature, dew point, relative humidity, and absolute moisture as used in ACDB. For these purposes, it is sufficient to interpolate air pressure from nearby sites. Wind speed and direction are critical variables and for sites where they are not measured data from a sufficiently comparable nearby station is substituted.



**Figure 1.** Locations of 82 reference sites. Colours denote the 69 NatHERS 2011 Climate Zones.

Solar radiation has been measured to the presently-accepted Bureau standards at only 19 sites. In the past, many sites and time periods had just daily totals of global horizontal irradiance,  $G$ , whether measured or estimated from satellite observations. NatHERS requires hourly values of  $G$ , diffuse horizontal irradiance,  $F$ , and direct normal irradiance,  $R$ , for all sites and times.

Since 2010, the Bureau of Meteorology has been able to provide radiation data derived from geostationary satellite data at 0.05° x 0.05° (approximately 5 km x 5 km) resolution for the entire continent. The dataset now extends from 1990 to the present. It is an exceptional resource, and now puts almost no limitation on the selection of climate zones and representative stations throughout the Australian land mass. Instead, the limitation now comes from the availability, completeness, and accuracy of data for the other required meteorological variables.

The 2019 revision of NatHERS was able to use the same sources as the 2016 revision for almost all of the 83 sites. As before, #31 WS: Willis Island, could not be updated, but its alternative, the new #83 XI: Christmas Island, had three years of new data. Because some Bureau stations have ceased recording one or more of the required data types, new Bureau stations were needed for #11 CH: Coffs Harbour, #22 SE: East Sale, #37 HA: Halls Creek, and #73 MN: Maleny.

The primary sites for each of the 83 climate zones are listed in Table 1. The column labelled 'Zone' gives an alternative numbering that corresponds to the eight major climate zones ('CZ') of the Australian Building Codes Board (ABCB), as shown in Figure 2. All sites have an intuitive two-letter code ('L2'), which also extends to 18 climate zones in New Zealand (Liley, J B, Sturman et al. 2008). Column 4 of Table 1, 'NH', gives the NatHERS numbers, which reflect when zones were created, at least within groups (1-29, 30-69, 70-80, 81-). Also listed are 'Post' code (in terms of which a NatHERS zone is assigned), location, 'Time' zone, Bureau of Meteorology and 'WMO' identifiers for the climate station, and the start year for data used herein for that site.

In Figure 2, locations of the NatHERS representative sites are shown with red circles. To avoid a clash with the existing labels on the ABCB plot, the locations are not labelled, but they can be understood by comparison with Figure 1 where the sites are named. In Figure 2, the interiors of the circles are coloured according to the relevant CZ. For exact agreement of the CZ assignments with the boundaries in the ABCB plot, developed from a map produced by the Bureau of Meteorology, the interiors of the circles should match the colour of their surrounds.

It is apparent that there is a mismatch for #47 BI: Bickley (CZ 4 but within zone 5); #51 FO: Forrest (CZ 5 but in zone 4); #65 OR: Orange (CZ 7 but in zone 4); and possibly #79 BL: Blue Mountains (CZ 6 but in zone 7). For Bickley, Orange, and Blue Mountains, the disagreement probably reflects the more detailed representation of climate regions within NatHERS, but for Forrest the reason is less clear. The ABCB map was updated in 2015, but the Climate Zone boundaries do not seem to have changed.

In 2013 we developed the new #81 BU: Busselton site, including a determination of which post codes it should contain (Liley, J Ben 2013). In 2016, we refined the technique to delineate the suggested boundary for #73 MN: Maleny, and to demonstrate the suitability of the new #83 XI: Christmas Island, in place of #31 WS: Willis Island (Liley, J Ben 2017). The technique could be deployed more generally to test the reliability of all the NatHERS zone boundaries, probably in conjunction with a move to describe the zones directly rather than as composed of post codes. The latter has the disadvantage that post codes are periodically revised for demographic or administrative reasons, and because of falling mail volumes, but there is no reason for them to reflect climatology. As noted in Liley (2013), the combination of 69 NatHERS zones and 83 sites means several zones contain more than one site, while several NatHERS zones have no site within their boundary and the reference site lies in another NatHERS zone and even ABCB zone (Liley, J Ben 2017, p 9).

No testing or redefinition of boundaries was undertaken for the present update.

**Table 1. Locations of representative sites for 83 NatHERS climate zones.** Column **Zone** follows ABCB zone coding, **CZ**, as illustrated in Figure 2; **L2** provides an additional intuitive zone identifier, **NH** is the NatHERS zone number; **Post** is the current post code; **Time** is the offset in hours from UT; **BoM** is the Bureau station number; **WMO** is the international code; **Since** is the start year for data from that site.

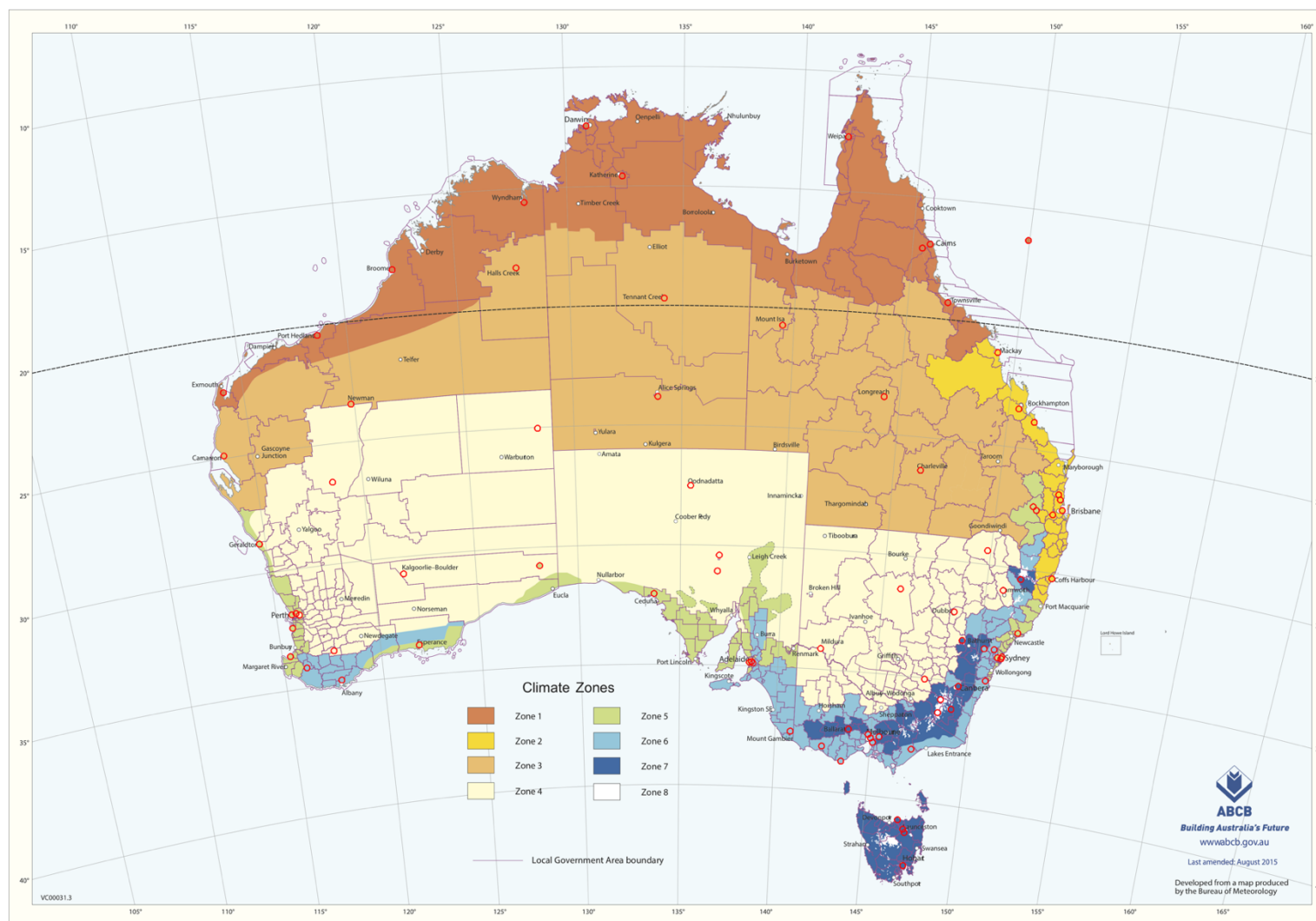
Zone	L2	Name	NH	Post	State	Altitude	Longitude	Latitude	Time	CZ	BoM	WMO	Since
CZ0101	DA	Darwin	1	820	NT	35.0	130.893	-12.424	9.5	1	14015	94120	1990
CZ0102	WP	Weipa	29	4874	Qld	19.0	141.921	-12.678	10.0	1	27045	94170	1990
CZ0103	KN	Katherine	74	853	NT	135.0	132.383	-14.523	9.5	1	14932	94131	1990
CZ0104	WY	Wyndham	30	6740	WA	4.3	128.150	-15.510	8.0	1	1006	95214	1990
CZ0105	WS	Willis Island	31	4871	Qld	9.8	149.965	-16.288	10.0	1	200283	94299	1977
CZ0106	CN	Cairns	32	4870	Qld	8.3	145.746	-16.874	10.0	1	31011	94287	1990
CZ0107	AT	Atherton	71	4880	Qld	473.1	145.428	-17.067	10.0	1	31210	94288	1990
CZ0108	BM	Broome	33	6725	WA	9.0	122.235	-17.948	8.0	1	3003	94203	1990
CZ0109	TO	Townsville	5	4814	Qld	9.1	146.766	-19.248	10.0	1	32040	94294	1990
CZ0110	HE	Pt Hedland	2	6721	WA	8.4	118.632	-20.372	8.0	1	4032	94312	1990
CZ0111	LM	Learmonth	34	6707	WA	5.5	114.097	-22.241	8.0	1	5007	94302	1990
CZ0112	XI	Christmas Island	83	6798	WA	4.0	96.834	-12.189	7.0	1	200284	96996	1995
CZ0201	MK	Mackay	35	4740	Qld	36.3	149.217	-21.117	10.0	2	33119	94367	1990
CZ0202	RO	Rockhampton	7	4700	Qld	15.1	150.477	-23.375	10.0	2	39083	94374	1990
CZ0203	GL	Gladstone	36	4680	Qld	75.2	151.263	-23.855	10.0	2	39123	94380	1990
CZ0204	MN	Maleny	73	4552	Qld	425.0	152.852	-26.753	10.0	2	40121	94547	2002
CZ0205	BR	Brisbane	10	4008	Qld	9.5	153.129	-27.392	10.0	2	40842	94578	1990
CZ0206	AM	Amberley	9	4306	Qld	24.9	152.711	-27.630	10.0	2	40004	94568	1990
CZ0207	CH	Coffs Harbour	11	2450	NSW	6.0	153.119	-30.311	10.0	2	59040	94791	1990
CZ0208	GM	Glasshouse Mountains	82	4519	Qld	48.0	152.962	-26.959	10.0	2	40284	95566	2002

Zone	L2	Name	NH	Post	State	Altitude	Longitude	Latitude	Time	CZ	BoM	WMO	Since
CZ0301	HA	Halls Creek	37	6770	WA	423.9	127.664	-18.229	8.0	3	2012	94212	1990
CZ0302	TE	Tennant Creek	38	872	NT	377.1	134.183	-19.642	9.5	3	15135	94238	1990
CZ0303	IS	Mt Isa	39	4825	Qld	341.0	139.488	-20.678	10.0	3	29127	94332	1990
CZ0304	LO	Longreach	3	4730	Qld	192.5	144.277	-23.437	10.0	3	36031	94346	1990
CZ0305	NE	Newman	40	6753	WA	524.5	119.799	-23.417	8.0	3	7176	94317	1990
CZ0306	AL	Alice Springs	6	872	NT	547.0	133.889	-23.795	9.5	3	15590	94326	1990
CZ0307	CR	Carnarvon	4	6701	WA	4.5	113.670	-24.888	8.0	3	6011	94300	1990
CZ0310	CV	Charleville	19	4470	Qld	303.3	146.256	-26.414	10.0	3	44021	94510	1990
CZ0401	GI	Giles	41	872	WA	599.0	128.301	-25.034	8.0	4	13017	94461	1990
CZ0402	MT	Meekatharra	42	6642	WA	519.0	118.537	-26.614	8.0	4	7045	94430	1990
CZ0403	OO	Oodnadatta	43	5734	SA	117.0	135.446	-27.555	9.5	4	17043	94476	1990
CZ0404	MO	Moree	8	2400	NSW	218.5	149.846	-29.491	10.0	4	53115	95527	1990
CZ0405	RX	Roxby Downs	72	5725	SA	99.7	136.877	-30.483	9.5	4	16096	95658	1998
CZ0406	KA	Kalgoorlie	44	6430	WA	366.0	121.453	-30.785	8.0	4	12038	94367	1990
CZ0407	TA	Tamworth	76	2340	NSW	395.9	150.836	-31.074	10.0	4	55325	95762	1990
CZ0408	WO	Woomera	45	5720	SA	168.5	136.805	-31.156	9.5	4	16001	94659	1990
CZ0409	CO	Cobar	46	2835	NSW	263.6	145.829	-31.484	10.0	4	48027	94711	1990
CZ0410	BI	Bickley	47	6076	WA	385.0	116.137	-32.007	8.0	4	9240	95610	1994
CZ0411	DU	Dubbo	48	2830	NSW	285.0	148.575	-32.221	10.0	4	65070	95719	1990
CZ0412	KT	Katanning	49	6317	WA	321.0	117.606	-33.686	8.0	4	10916	94641	1990
CZ0413	MI	Mildura	27	3500	Vic	51.1	142.087	-34.236	10.0	4	76031	94693	1990
CZ0414	WA	Wagga	20	2651	NSW	213.0	147.457	-35.158	10.0	4	72150	94910	1990

Zone	L2	Name	NH	Post	State	Altitude	Longitude	Latitude	Time	CZ	BoM	WMO	Since
CZ0501	OA	Oakey	50	4401	Qld	407.1	151.741	-27.403	10.0	5	41359	94552	1990
CZ0502	TW	Toowoomba	70	4350	Qld	641.5	151.913	-27.542	10.0	5	41529	95551	1990
CZ0503	GE	Geraldton	12	6532	WA	30.2	114.699	-28.805	8.0	5	8315	94403	1990
CZ0504	FO	Forrest	51	6434	WA	160.0	128.109	-30.845	8.0	5	11052	95646	1990
CZ0505	PE	Perth	13	6105	WA	20.0	115.976	-31.927	8.0	5	9021	94610	1990
CZ0506	SW	Swanbourne	52	6010	WA	41.0	115.762	-31.956	8.0	5	9215	94614	1994
CZ0507	CE	Ceduna	53	5690	SA	15.7	133.698	-32.130	9.5	5	18012	94653	1990
CZ0508	MD	Mandurah	54	6210	WA	3.5	115.712	-32.522	8.0	5	9977	94605	1990
CZ0509	WE	Williamstown	15	2318	NSW	7.9	151.836	-32.793	10.0	5	61078	94776	1990
CZ0510	EP	Esperance	55	6450	WA	27.0	121.893	-33.830	8.0	5	9789	94638	1990
CZ0511	PA	Parramatta	77	2200	NSW	7.5	150.986	-33.918	10.0	5	66137	94765	1990
CZ0512	SY	Sydney RO (Observatory Hill)	17	2000	NSW	40.2	151.205	-33.861	10.0	5	66062	94768	1990
CZ0513	MA	Mascot (Sydney Airport)	56	2020	NSW	5.0	151.173	-33.941	10.0	5	66037	94767	1990
CZ0514	AD	Adelaide	16	5067	SA	51.0	138.622	-34.921	9.5	5	23090	94675	1990
CZ0515	AC	Adelaide Coastal (AMO)	75	5950	SA	8.2	138.520	-34.952	9.5	5	23034	94672	1990
CZ0516	BU	Busselton	81	6280	WA	16.9	115.401	-33.686	8.0	5	9603	95611	1997
CZ0601	BL	Blue Mountains	79	2785	NSW	1080.0	150.274	-33.618	10.0	6	63292	94743	1990
CZ0602	RI	Richmond	28	2753	NSW	20.0	150.776	-33.600	10.0	6	67105	95753	1990
CZ0603	MJ	Manjimup	57	6258	WA	287.2	116.145	-34.251	8.0	6	9573	94617	1990
CZ0604	NO	Nowra	18	2540	NSW	105.0	150.535	-34.947	10.0	6	68072	94750	1990
CZ0605	AB	Albany	58	6330	WA	70.0	117.816	-34.941	8.0	6	9741	94802	1990
CZ0606	ML	Mt Lofty	59	5152	SA	685.0	138.709	-34.978	9.5	6	23842	95678	1990
CZ0607	TU	Tullamarine (Melbourne Airport)	60	3045	Vic	118.8	144.832	-37.666	10.0	6	86282	94866	1990



Zone	L2	Name	NH	Post	State	Altitude	Longitude	Latitude	Time	CZ	BoM	WMO	Since
CZ0608	CS	Coldstream	80	3770	Vic	83.9	145.409	-37.724	10.0	6	86383	94864	1995
CZ0609	ME	Melbourne RO	21	3053	Vic	32.2	144.970	-37.807	10.0	6	86071	94868	1990
CZ0610	MG	Mt Gambier	61	5291	SA	69.0	140.774	-37.747	9.5	6	26021	94821	1990
CZ0611	MR	Moorabbin	62	3194	Vic	12.7	145.096	-37.980	10.0	6	86077	94870	1990
CZ0612	SE	East Sale	22	3851	Vic	8.2	147.132	-38.116	10.0	6	85072	94907	1990
CZ0613	WR	Warrnambool	63	3275	Vic	71.4	142.452	-38.287	10.0	6	90186	94837	1990
CZ0614	OT	Cape Otway	64	3238	Vic	83.0	143.513	-38.856	10.0	6	90015	94842	1990
CZ0701	AA	Armidale	14	2350	NSW	1079.6	151.616	-30.527	10.0	7	56238	95773	1990
CZ0702	OR	Orange	65	2800	NSW	945.3	149.126	-33.377	10.0	7	63303	95726	1990
CZ0703	CA	Canberra	24	2609	ACT	580.0	149.201	-35.305	10.0	7	70014	94926	1990
CZ0704	SU	Sub-Alpine (Cooma Airport)	78	2630	NSW	931.0	148.973	-36.294	10.0	7	70217	94921	1991
CZ0705	BA	Ballarat	66	3355	Vic	435.6	143.791	-37.513	10.0	7	89002	94852	1990
CZ0706	LD	Low Head	67	7253	Tas	3.5	146.788	-41.055	10.0	7	91293	95964	1990
CZ0707	LT	Launceston (Ti Tree Bend)	23	7248	Tas	5.0	147.122	-41.419	10.0	7	91237	94969	1990
CZ0708	LU	Launceston Airport	68	7258	Tas	168.4	147.214	-41.549	10.0	7	91311	95966	1990
CZ0709	HO	Hobart	26	7004	Tas	51.4	147.328	-42.890	10.0	7	94029	94970	1990
CZ0801	CM	Cabramurra (old Alpine)	25	2720	NSW	1482.4	148.378	-35.937	10.0	8	72161	95916	1990
CZ0802	TH	Thredbo Village	69	2627	NSW	1380.0	148.304	-36.492	10.0	8	71041	95908	1990



**Figure 2.** Locations of 82 reference sites (red circles) within the ABCB Climate Zones. The interior of the circle shows the nominal Climate Zone for that site. Christmas Island (#83 XI), represented by Cocos (Keeling) Island at 12.2° S, 96.8° E, lies outside the map at upper left.

**Table 2. Additional Bureau stations for data missing from primary stations.** Data types listed alone, or marked with an asterisk, were the reason for using that station. If present, **Alt. P** is the barometer altitude. 'Cloud' indicates observations, 'Ceil.' denotes measurement by ceilometer.

Zone	L2	NH	BoM	BoM Station	State	Altitude	Alt. P	Latitude	Longitude							AWS
CZ0104	WY	30	1013	Wyndham	WA	11.0	16.0	-15.487	128.125							Cloud
CZ0107	AT	71	31108	Walkamin Research Station	Qld	594.0		-17.135	145.428							Cloud
CZ0203	GL	36	39326	Gladstone Airport	Qld	16.6	16.9	-23.870	151.221	Wind	T	T <sub>dew</sub>	P	Ceil.*		Y
CZ0204	MN	73	40651	Jimna Forestry	Qld	523.0		-26.664	152.461	Wind	T	T <sub>dew</sub>				Y
			40988	Nambour Daff Hillside	Qld	53.2	54.0	-26.644	152.938	Wind	T	T <sub>dew</sub>	P			Y
CZ0207	CH	11	59151	Coffs Harbour Airport	NSW	3.5	4.0	-30.319	153.116	Wind	T	T <sub>dew</sub>	P	Ceil.		Y
CZ0301	HA	37	2079	Halls Creek Airport	WA	409.4	410.7	-18.234	127.667	Wind	T	T <sub>dew</sub>	P	Ceil.		Y
CZ0405	RX	72	16065	Andamooka	SA	76.0		-30.449	137.169							Cloud
CZ0410	BI	47	9021	Perth Airport	WA	15.4	20.0	-31.927	115.976	Wind	T	T <sub>dew</sub>	P	Ceil.*		Y
CZ0412	KT	49	10579	Katanning Comparison	WA	310.0	311.0	-33.689	117.555							Cloud
CZ0503	GE	12	8051	Geraldton Airport Comparison	WA	33.0	35.0	-28.795	114.698	Wind	T	T <sub>dew</sub>	P	Ceil.		Y
CZ0506	SW	52	9172	Jandakot Aero	WA	30.0	30.7	-32.101	115.879	Wind	T	T <sub>dew</sub>	P	Ceil.*		Y
CZ0508	MD	54	9887	Mandurah	WA	21.0	22.0	-32.521	115.750	Wind	T	T <sub>dew</sub>	P			Y
			9194	Medina Research Centre	WA	14.0		-32.221	115.808							Cloud
CZ0511	PA	77	66137	Parramatta North (Masons Dr)	NSW	55.0		-33.792	151.018	Wind	T	T <sub>dew</sub>		Cloud*		
CZ0513	SY	17	66022	Fort Denison	NSW	2.0		-33.855	151.225	Wind						Y
CZ0514	AD	16	23034	Adelaide Airport	SA	2.0	8.2	-34.952	138.520	Wind	T	T <sub>dew</sub>	P	Ceil.*		Y
CZ0516	BU	81	9569	Busselton	WA	3.9		-33.655	115.319		T	T <sub>dew</sub>				Y

Zone	L2	NH	BoM	BoM Station	State	Altitude	Alt. P	Latitude	Longitude							AWS
CZ0601	BL	79	63039	Katoomba (Murri St)	NSW	1015.0		-33.712	150.309						Cloud	
CZ0603	MJ	57	9592	Pemberton	WA	174.0	175.0	-34.448	116.043						Cloud	
			9510	Bridgetown Comparison	WA	149.9	150.7	-33.957	116.137						Cloud	
CZ0605	AB	58	9741	Albany Airport Comparison	WA	68.0	69.0	-34.941	117.802	Wind	T	T <sub>dew</sub>	P	Ceil.	Y	
CZ0606	ML	59	23878	Mount Crawford	SA	525.0	525.5	-34.725	138.928	Wind	T	T <sub>dew</sub>	P		Y	
			23733	Mount Barker	SA	363.0		-35.064	138.851						Cloud	
CZ0608	CS	80	86071	Melbourne Regional Office	Vic	31.1	32.2	-37.807	144.970	Wind	T	T <sub>dew</sub>	P	Cloud*	Y	
CZ0609	ME	21	86068	Viewbank	Vic	66.1	66.4	-37.741	145.097	Wind	T	T <sub>dew</sub>	P		Y	
			86338	Melbourne (Olympic Park)	Vic	7.5	7.5	-37.826	144.982	Wind	T	T <sub>dew</sub>	P		Y	
CZ0612	SE	22	85314	East Sale Airport	Vic	5.0	8.0	-38.102	147.140	Wind	T	T <sub>dew</sub>	P	Ceil	Y	
CZ0613	WR	63	90171	Cashmore Airport	Vic	80.9	81.5	-38.315	141.471	Wind	T	T <sub>dew</sub>	P	Ceil.*	Y	
CZ0701	AA	14	56037	Armidale (Tree Group Nursery)	NSW	987.0		-30.524	151.672						Cloud	
CZ0702	OR	65	63231	Orange Airport Comparison	NSW	948.0		-33.381	149.123	Wind	T	T <sub>dew</sub>		Cloud*		
			63303	Orange Airport	NSW	944.7	945.3	-33.377	149.126	Wind	T	T <sub>dew</sub>	P	Ceil.	Y	
CZ0703	CA	24	70351	Canberra Airport	NSW	577.1	577.6	-35.309	149.200	Wind	T	T <sub>dew</sub>		Ceil.	Y	
CZ0705	BA	66	89105	Lookout Hill	Vic	965.0		-37.282	143.247	Wind	T	T <sub>dew</sub>			Y	
CZ0706	LD	67	91126	Devonport Airport	Tas	8.0	9.5	-41.170	146.429	Wind	T*	T <sub>dew</sub>	P	Ceil.	Y	
CZ0707	LT	23	91311	Launceston Airport	Tas	166.9	168.4	-41.549	147.214	Wind	T	T <sub>dew</sub>	P	Ceil.*	Y	
CZ0708	LU	68	91104	Launceston Airport Comp.	Tas	166.0	178.0	-41.450	147.203	Wind	T	T <sub>dew</sub>	P	Ceil.	Y	
CZ0709	HO	26	94008	Hobart Airport	Tas	4.0	27.4	-42.834	147.503	Wind	T	T <sub>dew</sub>	P	Ceil.*	Y	
CZ0801	CM	25	72043	Tumbarumba Post Office	NSW	645.0		-35.778	148.012						Cloud	
CZ0802	TH	69	71032	Thredbo AWS	NSW	1957.0	1367.9	-36.492	148.286	Wind*	T	T <sub>dew</sub>	P		Y	

## 2.2 Additional sites

As previously, additional sites are needed to complete the time series for many primary locations, either because of a missing weather element or because of discontinued measurements. The extra sites are listed in Table 2, which also indicates which weather element necessitated their use.

The majority of climate stations used for NatHERS, both primary and additional, are Automatic Weather Stations (AWS) in the Bureau of Meteorology network. The data are supplied as half-hourly records, though sometimes recording is hourly, so alternate values are missing. In 2016, a request to the Bureau for all available AWS data since 1990 yielded values for 665 AWS sites. The corresponding 2019 request for data from 2016 to the present yielded values for 631 sites. Of the latter, 66 were new stations, sometimes as replacements for older stations.

The 2019 revision of NatHERS used hourly data from 104 AWSs, plus values from synoptic data (three-, six-, or twelve-hourly) for 16 staffed sites, interpolated to hourly steps. As in 2016, the 2019 revision did not use data before in 1990, except by retaining #31 WS: Willis Island unchanged.

Table 2 includes all additional sites used in both the 2016 and current revisions, though some ceased recording before the start of new data on 1 January 2016. The primary sites listed in Table 1 are unchanged from the 2016 revision.

As can be discerned by comparison of Table 2 with the corresponding table in Liley (2017), four new AWS data series were needed for this revision:

- BoM #40988, Nambour Daff Hillside, for #73 MN: Maleny
- BoM #59151, Coffs Harbour Airport, for #11 CH: Coffs Harbour
- BoM #02079, Halls Creek Airport, for #37 HA: Halls Creek
- BoM #85314, East Sale Airport, for #22 SE: East Sale.

As the station names suggest, the latter three of these were obvious substitutions of nearly collocated stations, but the first is not. NatHERS zone #73 MN: Maleny, is represented by a synoptic station BoM #40121, Tamarind St, Maleny, at 425 metres above sea level. Wind data were obtained from AWS #40651, Jimna Forestry, at 523 m asl and 40 km to the west. Hourly air temperatures and dew point from the AWS were also used with kriging to generate the fitted surface by which synoptic data are interpolated to hourly estimates. However, AWS #40651 discontinued those measurements in September 2017, so AWS #40988, as above, has been used instead. It is just 15 km NE of Maleny, but as it is only 54 m asl there was a question as to whether it adequately represented the higher and cooler inland climate of Maleny. Comparison of the #40988 wind data with simultaneous synoptic values from Tamarind St showed a similar pattern of direction and speed. Temperatures at Nambour Daff were closely correlated and around 3 °C higher, consistent with the altitude difference. Dew point temperatures were strongly correlated, with minimal bias.

## 2.3 Cloud data

As shown in Table 2, many stations were selected for their cloud data. Of the AWS's that were primary for their site, 53 had ceilometers, which measure the time average of the base height of overhead cloud below 3,700 m altitude to estimate cloud cover. Cloud observations are a much better source of this information for calculations of radiative energy balance as used in building

simulation, but their frequency is usually much lower. Where available, cloud cover records from human observers were used.

During daytime hours for the period of satellite-derived irradiance, described below, those data were used to estimate cloud cover where other measures were unavailable. The total satellite-derived global irradiance over the five  $0.05^\circ \times 0.05^\circ$  ( $\sim 5$  km square) pixels nearest to the target site, expressed as a fraction of the equivalent clear sky irradiance, gave a measure of fractional cloud cover that correlated well with observations where they were available.

## 2.4 Wind data

The 2016 update of NatHERS data included a substantial review of wind data. Plots like those in Appendix A showed marked changes in median wind speeds at certain times, and these were found to coincide with records of site visits and instrument changes. Over time, many anemometers were moved from near ground (typically 2 m) to 8-m masts, as assumed for NatHERS data. In other instances, there were gaps in data records that had to be filled from a nearby climate station.

Changes in wind regime are not confined to remote areas where few houses would be affected. Wind speeds are highly variable in built-up areas, and they are strongly damped by trees, so the problem is even greater where population density is high. In the 2016 revision, new data sources were needed for several sites, including Melbourne, and the latter was discussed in the most recent report (Liley, Ben 2019). In the current revision, the only new source of wind data was BoM #40988, Nambour Daff Hillside, as described above.

## 3 Data quality control

### 3.1 Error detection

Past updates of NatHERS data have entailed the development of new algorithms for quality-control (QC) to ensure errors are treated consistently and objectively, and those algorithms were detailed in the respective reports (Liley, J Ben 2013; Liley, J Ben 2017). The retained time series from the 2016 revision already had errors flagged, so QC was only needed for the new data. No new QC algorithms were found to be necessary for that but, as described below, some further anomalies were found in earlier data.

For completeness of this report, we include the 2016 description of QC procedures.

### 3.2 Missing data

Temperature, dew point, and mean sea level pressure (*MSLP*) from all measurements, and from temporal interpolation of synoptic data, are combined for spatial interpolation by kriging to fill any missing values. It is necessary to interpolate synoptic measurements to hourly values before kriging; otherwise the many more measurements at 0900, 1200, 1500, etc., than at adjacent times affects the relative weightings, resulting in unphysical temporal steps in the resulting time series.

Because *MSLP* varies smoothly in both space and time, station pressure can be reliably derived from interpolated *MSLP* even for sites with no measurements of pressure. Spatially interpolated temperature and dew point are used in the RMYs if it is unavoidable, but months with measured hourly or synoptic values are preferred in the selection of representative years.

### 3.3 Visible anomalies

Plots of climate time series, as illustrated in Appendix A with corrected data, were generated for all 83 sites and reviewed visually. Such inspection is the best method for finding extended periods of anomalous data, but it is then necessary to find algorithms that can be applied objectively to detect all instances. This process is inevitably iterative, ensuring tolerable levels of non-detection versus false detection, and assessing the effect on other processes.

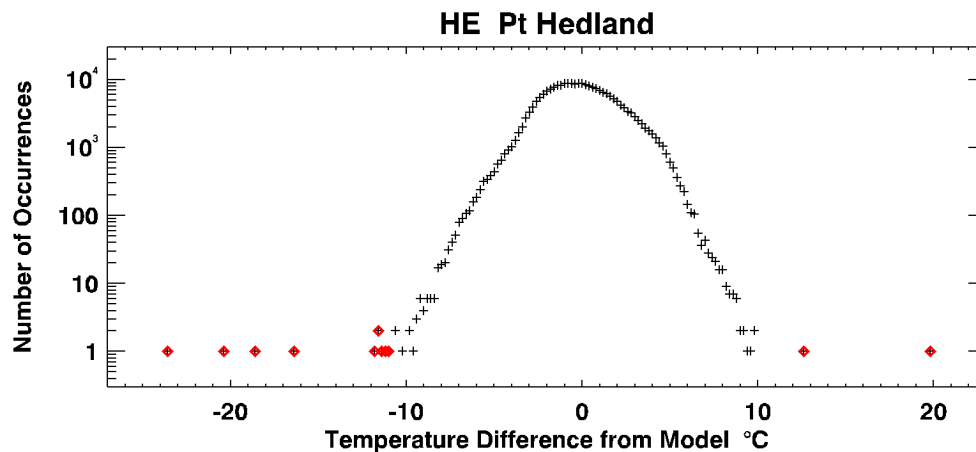
Even in extensive review, especially of full time series like those in Appendix A, individual anomalies may not stand out. It is common for climate data series to contain occasional peaks or zero values from instrument malfunction, or from site visits and instrument servicing. Ideally the erroneous data will have been flagged by the Bureau of Meteorology with data quality indicators, but even without such flags many errors of this type can be found by a series of techniques.

### 3.4 Statistical anomalies

For most meteorological or climate variables, it is possible to set limits beyond which any measurement is suspect. These limits should vary with site, and often with season, and they can be chosen from review of the time series. However many odd values do not stand out except by comparison with adjacent values. We have found good discrimination of anomalies with the following technique, applied here to temperature, pressure, absolute moisture, and wind speed.

A histogram of a sufficient number of values in a random normal time series would show the classic bell-curve shape with  $y \propto \exp(-x^2)$ . With a logarithmic  $y$  axis, the curve is a negative quadratic. Real climatic time series have underlying structure, together with seemingly random processes on

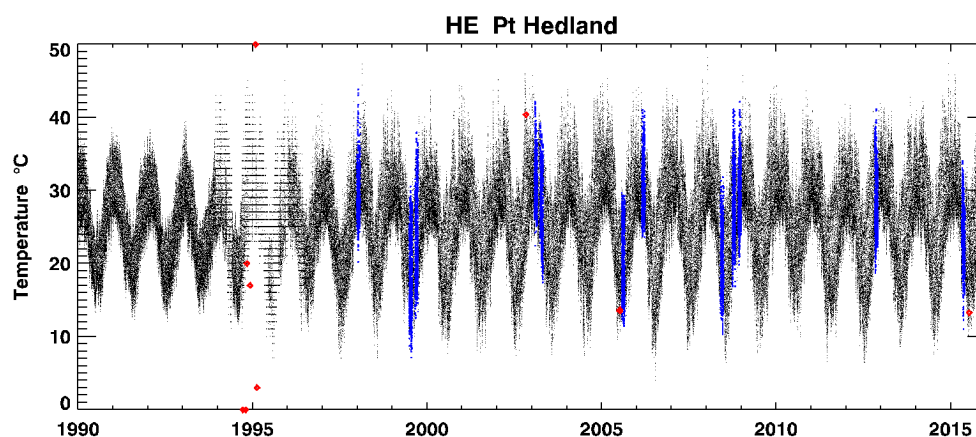
many time scales. For temperature, illustrated here for Port Hedland, we fitted a simple model of the seasonally varying diurnal cycle, subtracted it from the data, and smoothed them to a timescale typical for weather systems. Individual temperature differences from the result are represented in Figure 3 as a histogram with logarithmic ordinate. Points outside a fitted envelope are highlighted.



**Figure 3.** Histogram of hourly temperature differences from mean-smoothed values, after subtraction of fitted seasonal and diurnal cycles. Values highlighted in red are likely to be anomalous.

Several of the values highlighted in red clearly depart from the overall pattern, even allowing for the ‘digitisation’ error of presenting integer values on a logarithmic scale. Another view of this analysis is that progressively lower values on the y axis correspond to greater rarity; with 228,000 data values, a horizontal line at  $N = 23$  would demarcate ‘one in 10,000’ occurrence rates, and this could be the basis for excluding certain data. Instead, we observe that progressively larger data sets will include rarer but still genuine events, whereas points outside the pyramidal envelope seem incongruous. Here, six of the differences from adjacent values are obviously anomalous. There are five points between -12 and -10 °C, selected by the algorithm, that are believable on close inspection of the time series (not shown), but similar instances for other sites are not. We include them here to illustrate the difficulty in finding objective algorithms with low rates of false positives or false negatives.

Figure 4 shows the full time series, with the anomalous values again highlighted in red.

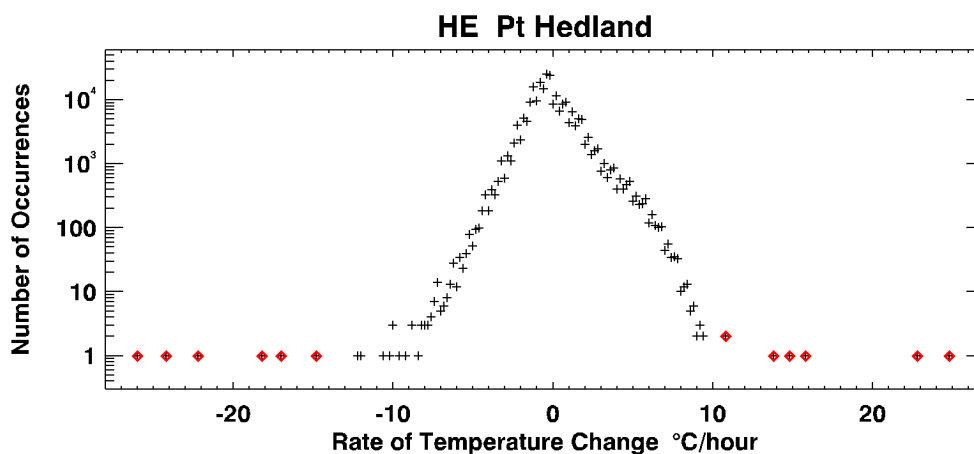


**Figure 4.** Time series of hourly temperature, with RMY data highlighted in blue. Eleven values, highlighted in red, were found by the outlier algorithm to be anomalous.



Figure 4 also shows (in blue) the year-months selected for the RMY. Although none of the red points were in the (subsequently) selected data, the inclusion or exclusion of anomalous values still affects the statistical procedure for TMY month selection described later.

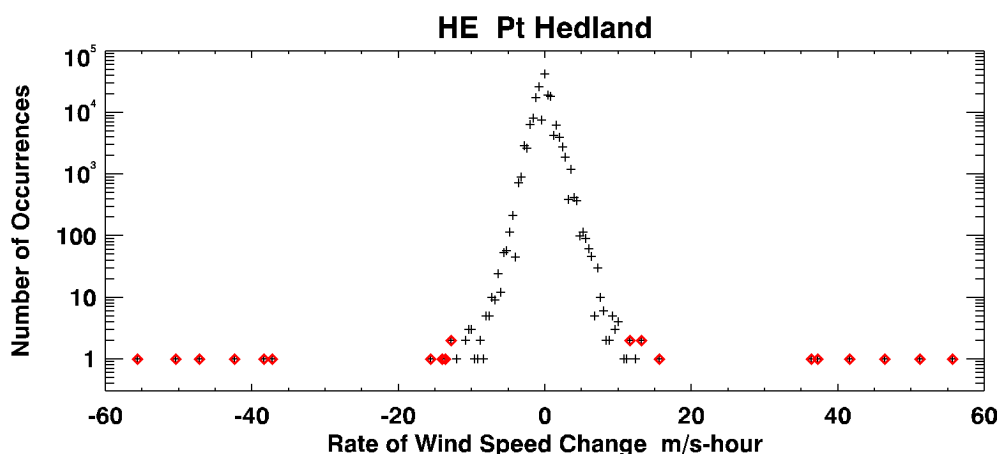
These anomalies are mostly just single hourly values. Sequences of two or more anomalous values are detected by an algorithm that uses the histogram of all hour-to-hour differences, as in Figure 5.



**Figure 5.** Histogram of hour-to-hour temperature differences. Values highlighted in red are data points removed by the two algorithms.

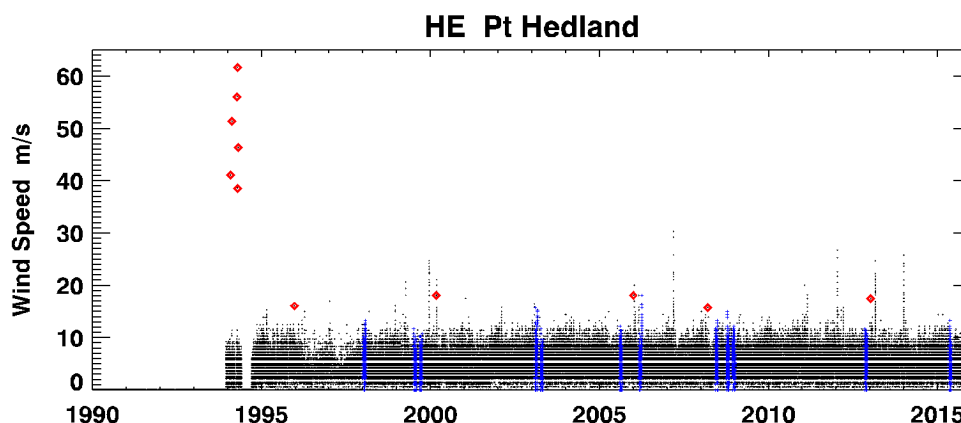
For this analysis, which detects instances of excessive rate of change in a variable, it is no longer necessary to first remove any seasonal cycle. Some errors are step changes in values, as can occur when an instrument malfunctions for a short period, or from transcription errors in manual records. The algorithm examines instances where two steps of opposite sign occur in temporal proximity, suggesting that values in the interval between them are incorrect.

Wind data for northern, especially coastal, locations like Port Hedland include incidents of gale force winds. For this reason, it is difficult to establish limits in absolute value, but rate of change in wind speed is still informative. The histogram of wind speed changes from hour to hour at Port Hedland is shown in Figure 6, analogous to Figure 5 for temperatures. Red diamonds again denote atypical values of the differences according to the algorithm.



**Figure 6.** Histogram of hourly wind speed change from previous value. The values highlighted in red, detected by algorithm, are considered anomalous.

Visual inspection of these 18 values confirmed that all are indeed dubious. Figure 7 shows the corresponding time series, analogous to Figure 4. Reassuringly, we can see that the detection algorithm has not rejected the periods of high winds that rise and fall with reasonable continuity, typically in connection with storm systems.



**Figure 7.** Time series of hourly wind speed, with RMY data highlighted in blue. The values highlighted in red are anomalous, and the most extreme values are discounted, but regular incidents of high winds pass the test.

With the above criteria applied objectively to the large volume of data, it is not possible to explore each detection forensically, but many instances are reviewed to strike a balance between correct and false detection of errors. Good quality control algorithms should detect the great majority of erroneous data but accept genuine weather extremes. All the algorithms described here were developed and refined iteratively across the datasets for all sites, in order to make the process as objectively reproducible as possible.

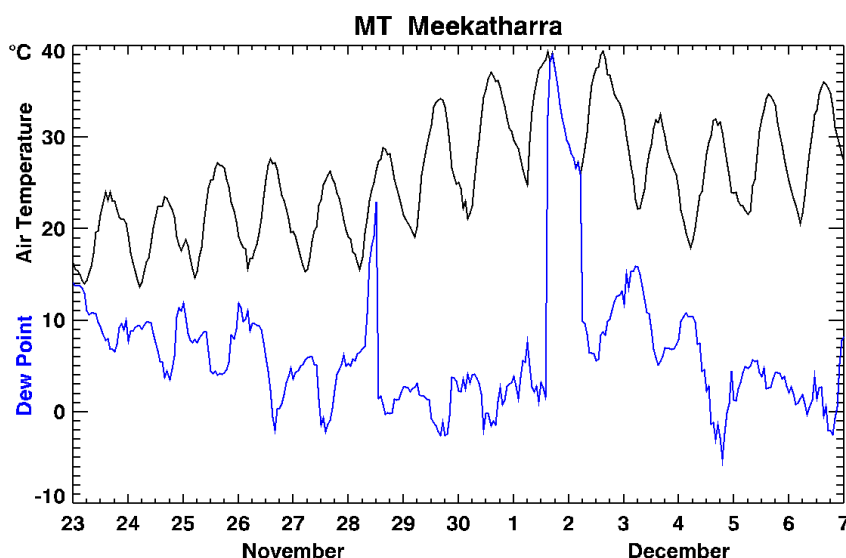
### 3.5 Tropical cyclones

The same analysis for pressure did not always avoid false positives, because pressure changes slowly other than in exceptional circumstances. A classic example from our original analysis of ACDB data for Darwin found the pressure drop, by 30 hPa over three hours, early on 25 December 1974 exceeded all plausible limits for the detection algorithm. The data would thus be marked as erroneous. As history records, they were very real, from the passage of cyclone Tracy.

Even such extremes in pressure data are probably unimportant for NatHERS in routine use, as the effect on moisture calculations is minor. On the other hand, tropical cyclones at northern latitudes are an important design consideration for the Australian building industry, and the NatHERS time series might need to be considered in this context. Thus, we check any comparable drops in atmospheric pressure against a Bureau of Meteorology database of tropical cyclone tracks, retaining any incidents within 150 km and 6 hours of the storm passage.

### 3.6 Drying of wet bulb wick

A major source of errors in past data appeared as extreme values of absolute humidity at dry inland sites. The problem permeated both the 2008 revision of ACDB (Energy Partners 2008) and raw Bureau datasets. Most AWS humidity data come from electronic sensors, but they are less accurate than traditional wet bulb measurements in very dry conditions. However, it is in just those conditions that wet bulb measurements are compromised if the reservoir is not refilled in time. Such an occurrence is illustrated in Figure 8 for Meekatharra, 430 km from the coast in Western Australia.



**Figure 8.** Hourly data from the Bureau of Meteorology, AWS site 7045, in 2008. Calculated absolute moisture is nearly 50 g kg<sup>-1</sup> when air temperature approaches 40 °C at 100% relative humidity.

On the afternoon of 1 December 2008, the absolute moisture content calculated from temperature and dew point reaches 50 grams of water vapour per kilogram of air, more than twice credible extremes. Dew points above 33 °C are extremely rare internationally, and are at or beyond the limit of human survivability as the body loses its ability to remove heat by evaporation. Instead, what has occurred is that the wick of the wet bulb thermometer has dried, so that the ostensible ‘wet bulb’ temperature has risen sharply to meet the dry bulb temperature. That the dry bulb temperature follows the same diurnal cycle as on adjacent days of low humidity is a sure sign that the dew point data on that afternoon, and similarly on the morning of 28 November, should be discounted.

This problem was prevalent in past ACDB data, and it is common even in recent Bureau data; perhaps more so if sites are visited less frequently. The algorithm that we developed to detect it uses a combination of criteria tested across a large number of data sets. From analysis of the distribution of dew points for all sites, we determined credible peak values as a function of latitude and altitude:

$$T_{dx} = 32.2 \text{ (}^{\circ}\text{C)} - 0.13 \text{ (}^{\circ}\text{C/}^{\circ}\text{S)} \times \text{Latitude} - 0.004 \text{ (}^{\circ}\text{C/masl)} \times \text{Altitude}$$

Values range from 21.6 to 30.6 °C over the range of NatHERS sites. The filter rejects dew point values above the value for that site, or where wet bulb temperature matches dry bulb and values change too rapidly. From wet bulb data that are not suspect, the limits on accepted rate of change are an increase of 1.5 °C per hour, or a decrease of -2 °C per hour. Rejected values are marked as missing, to be filled by interpolation like other gaps in dew point data.

### 3.7 Wind data

As noted above, the 2016 update involved a thorough review of wind mast and anemometer changes, based on Bureau of Meteorology station metadata reports, as described at <http://www.bom.gov.au/climate/data/stations/about-weather-station-data.shtml#metadata>

Whether from different instrumentation, an obviously different wind regime at the same site, or a new site with different wind climate, discontinuities in wind are a problem for RMY selection. The statistical procedure described in Section 5 copes sensibly with pseudo-random variation within a consistent distribution, but not with substantial shift in the distribution over time. Instead, it is necessary to determine which distribution best represents the site.

We addressed this issue in 2016, for any obviously different earlier periods at the same site, or from adjacent sites with different wind climate. Provided the distributions of wind directions were similar, the less representative wind speeds were transformed for reasonable consistency of the median (50<sup>th</sup>), 90<sup>th</sup>, and 98<sup>th</sup> percentiles with those of good data. The results are illustrated in the plots of Appendix A, which show those percentiles.

Reliance on the transform is not ideal, because low wind speed at 8 m height would often mean no detectable wind was measured at 2 m, or at a more sheltered site. Better results may be achievable with a more extensive analysis of data from nearby anemometers, as undertaken in 2016 for #21 ME: Melbourne RO, and described in Liley (2019). That would be a large amount of work, but add little value relative to the differences in wind conditions within a NatHERS climate zone.

## 4 Solar radiation

### 4.1 Ground-based data

The measures of solar radiation used in NatHERS are global horizontal irradiance,  $G$ , diffuse horizontal irradiance,  $F$ , and direct normal irradiance,  $R$ . The three quantities are linked by the relationship  $G = F + R \cdot \cos(Z)$ , where  $Z$  is the solar zenith angle. Measurement of  $G$  is straightforward, using fixed pyranometers that require little maintenance other than cleaning but need periodic recalibration. By use of a shade band that moves with season, an additional instrument of the same type provides a measure that can be used to estimate  $F$ . Measurement of  $R$  requires a pyrheliometer with narrow field of view that accurately tracks the sun, and usually the same device is used with a shade disk to provide an accurate measure of  $F$  by blocking less of the sky than the shade band does.

Prior to the 1990s, many stations in the Bureau of Meteorology network measured  $G$ , and often  $F$ , with the former type of instrument. Such data are usually of variable quality, and in the 1990s the Bureau changed to using systems of the latter type, which can provide data of much higher quality but are more labour-intensive. There are now datasets of this type for 20 Bureau climate stations. Nineteen of those stations are representative sites in NatHERS (only Cape Grim, on the western tip of Tasmania, is not). The measurements are made to the standards of the Baseline Surface Radiation Network (BSRN), used globally for validating models of radiative energy balance and for ground-truth of radiation data derived from satellite-based instruments.

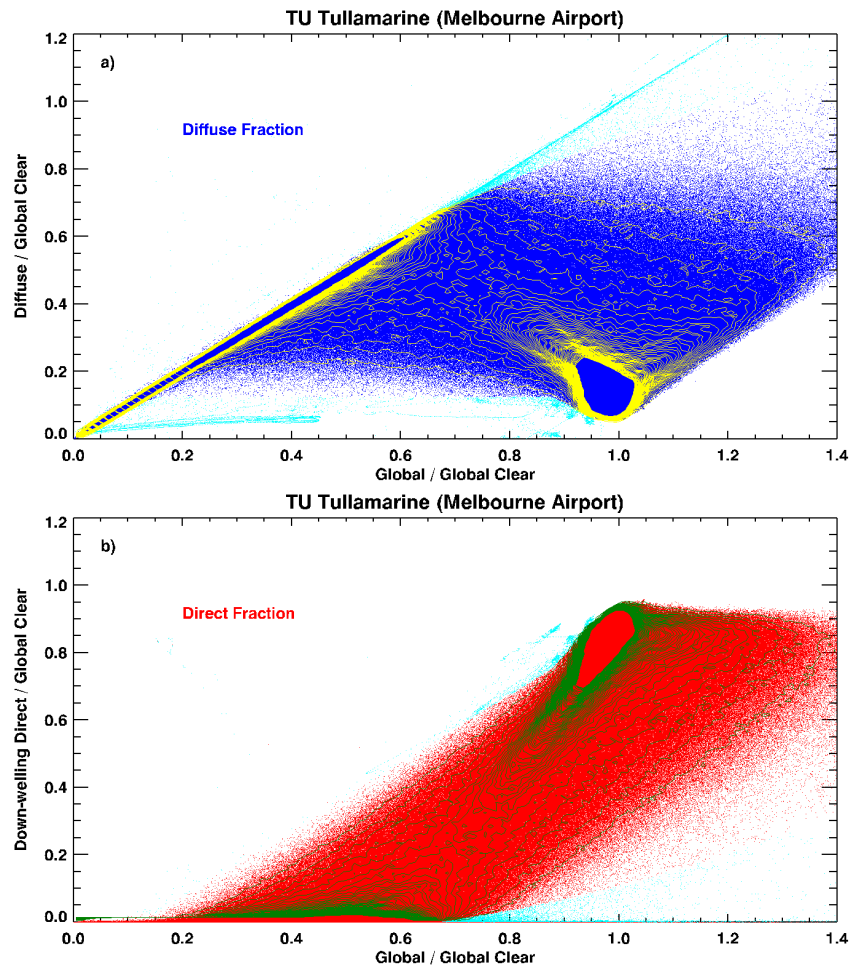
In clear sky conditions, all three of  $G$ ,  $F$ , and  $R$  can be modelled with good precision by radiative transfer models. The main uncertainty relates to aerosol optical depth (AOD), high values of which are associated with low meteorological ‘visibility’ or ‘visual range’. There is also some dependence on atmospheric water vapour and ozone, but generally less than for aerosol.

In the relatively clean conditions of Australia, other than when affected by dust clouds or smoke, aerosol effects are small. Measurements of the components of radiation in cloud-free conditions are reasonably well described by simple functional relationships that depend only on solar zenith angle  $Z$ , and a simple seasonal cycle in AOD for the model of  $R$ .

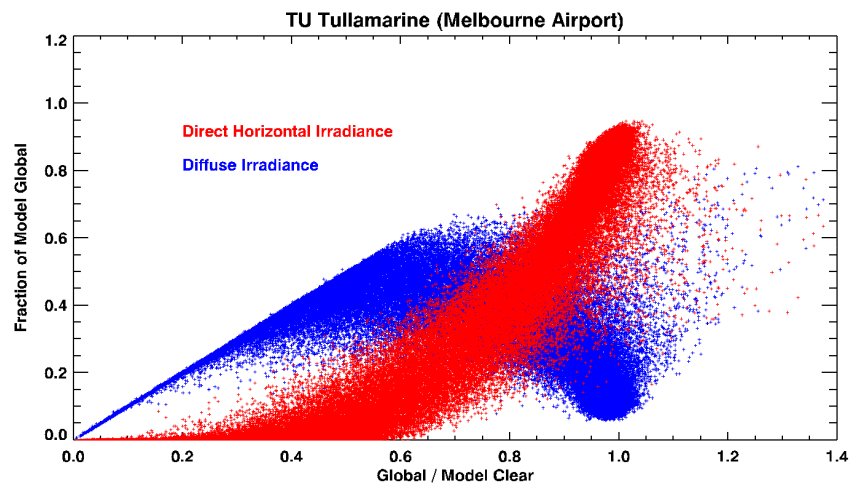
In past work, we have found most use for the model clear-sky global irradiance,  $G_c$ , simply parameterised as  $G_c = G_0 \cdot \cos(Z)^b$ . Both  $G_0$  and  $b$  are evaluated by fitting to measurements on clear sky days, and we find values of  $G_0 \approx 1120 \text{ W m}^{-2}$  and  $b \approx 1.14$  provide a good fit across all the Australian radiation sites. Conceptually,  $G_0$  represents the irradiance for overhead sun ( $Z = 0$ ).

As illustrated in previous reports, and here in Figure 9, the diffuse and down-welling direct components of surface irradiance show compact distributions when expressed relative to  $G_c$ . Though the scattergrams, each containing 3.7 million data points, appear to fill a broad range, the contour plot shows that the great majority of points fall in two groups. For measured  $G$  less than about half of  $G_c$ , the sun is obscured by cloud, so  $R$  is near zero, and irradiance is all diffuse,  $F = G$ . Under clear sky,  $G = G_c$ , irradiance is predominantly down-welling direct,  $R \cdot \cos(Z) \approx 0.9 \cdot G$ , with little diffuse,  $F \approx 0.1 \cdot G$ .

The contours show that intermediate points between these regimes are infrequent in 1-minute data, but accumulating to hourly values fills in the distributions for  $0.5 \cdot G_c < G < G_c$ , as shown in Figure 10. These relationships provide a means for estimating  $F$  and  $R$  if only  $G$  is known, and  $G_c$  is calculated as above. They also provide tests of validity for quality control of radiation values, as discussed in previous reports (Liley, J Ben 2013; Liley, J Ben 2017). For example, the points coloured light blue in Figure 9 are rejected as anomalous.



**Figure 9.** Measured a) diffuse (blue), and b) downward direct (red) components of global irradiance, all as fractions of model clear sky global irradiance. Points are 1-minute data for Melbourne Airport that satisfy Bureau criteria and  $Z < 85^\circ$ . Contours show the great majority of points are in two clusters. Points coloured light blue are anomalous, typically because of misalignment.

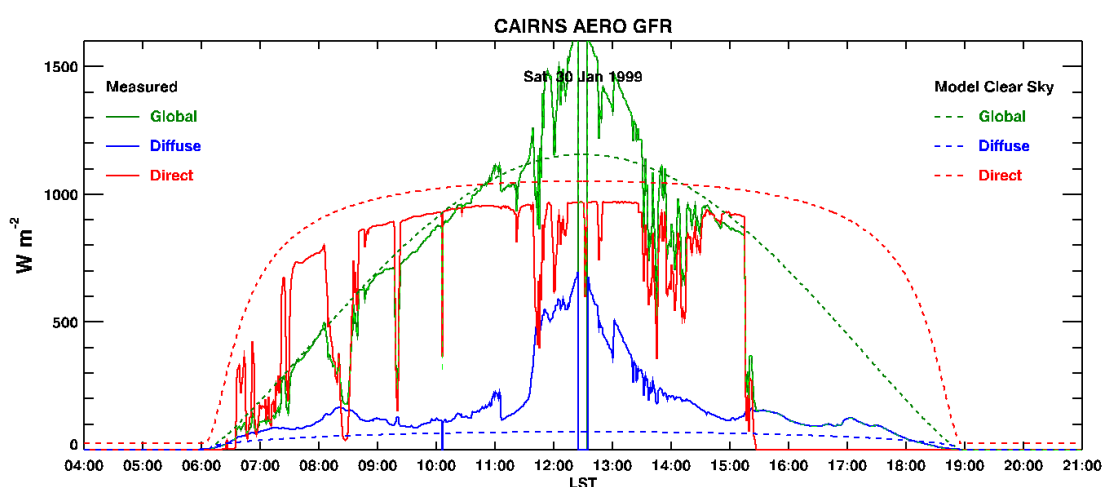


**Figure 10.** Hourly averages of diffuse (blue) and down-welling direct (red) components of global irradiance. Both abscissa and ordinate are again expressed as The scattergram shows hourly averages at Melbourne Airport for  $Z < 85^\circ$ .

The points in Figure 9 to the right of the clear sky clusters show instances of ‘cloud enhancement’, a familiar phenomenon where global irradiance is greater than it would be under clear sky ( $G > G_c$ ) because the sun is unobscured and bright cloud gives much more diffuse light than does blue sky. The phenomenon is common in 1-minute data for scattered or broken cloud conditions, even to enhancements of 50% ( $G > 1.5 \cdot G_c$ ). It mostly disappears from hourly averages (Figure 10) because it is very rare for the sun to be unobscured for an hour in the presence of extensive bright cloud.

## 4.2 Errors in ground-based radiation data

In the course of the 2019 revision, we found that enhancement above 50% did occur in hourly aggregates of data for #32 CN: Cairns (BoM #31011) in January 1999, and this month is used in the RMYs for Cairns in both the 2016 and initial 2019 revisions.



**Figure 11.** Diffuse (blue) and direct (red) components of global irradiance (green) measured (solid) and modelled (dashed) at Cairns on 30 January 1999. The high values of global and diffuse around noon are incongruous.

Inspection of the data from this period showed patterns like that in Figure 11 from 30 January 1999. This was a mystery, as it would require a very bright halo of cloud surrounding the sun almost continuously for two hours, or else substantial reflection into both the global and diffuse sensors in suspiciously equal amounts. We did not initially suspect the more obvious interpretation that the shade disk for the diffuse sensor was misaligned, for two reasons. First, we understood that the radiation data were subjected to quality control that would detect such errors. Second, the problem should not affect the measurements of  $G$ , so it should be immediately apparent as a violation of the consistency relationship  $G = F + R \cdot \cos(Z)$ .

From very helpful discussion with Dr Bruce Forgan, who led the Bureau of Meteorology’s radiation research for more than three decades before his retirement, both points were answered. Standards of instrument maintenance (especially alignment and cleaning) had lapsed at times at several sites, and errors were not always detected in subsequent review. That should still have been detectable from the consistency check, except for our misunderstanding of the data as supplied.

When both  $F$  and  $R$  are measured reliably, the ‘component sum’  $G_0 = F + R \cdot \cos(Z)$  is a better measure of  $G$  than that obtained directly from the global pyranometer. Indeed, it is for this reason the BSRN standards require the measurement of all the separate components  $F$  and  $R$  rather than just  $G$ . The



utility of the separate components for calculating radiant flux onto angled surfaces, as in NatHERS, is just an additional benefit.

Because of this hierarchy, the files from the Bureau of Meteorology containing the 1-minute data actually report  $G_0$  as  $G$  when data have passed quality control. In consequence, all the data automatically passed our checks of consistency.

With this realisation, we reappraised all of the 1-minute radiation data, using data filters to check their consistency with the relationships illustrated in Figure 9. The search found 164 periods, from a few hours to many days in length, when data were invalid. These periods were masked out in the aggregation of 1-minute data to hourly values, and the RMY files for radiation sites recalculated for both 2016 and 2019 NatHERS revisions.

### 4.3 Satellite-derived radiation data

As used in NatHERS 2012 and 2016, and described in the related reports, research by the Bureau of Meteorology (Grant 2009) has provided hourly solar radiation data anywhere on the Australian continent. In the initial release, estimates of global and direct irradiance for each hour from geostationary satellite data cover the period 1 January 1998 to 31 December 2010, excluding 1 July 2001 to 30 June 2003. A subsequent release of these data extended the coverage to the full period from 1 January 1990 to 31 December 2012, with small changes resulting from comparison with ground-based data. For the NatHERS 2016 update, the data extension to the end of 2015 used the same sources and processing.

The data are derived from satellite imagery processed by the Bureau of Meteorology from the Geostationary Meteorological Satellite and MTSAT series operated by Japan Meteorological Agency and from GOES-9 operated by the National Oceanographic & Atmospheric Administration (NOAA) for the Japan Meteorological Agency. Since March 2016, data are derived from imagery acquired by Himawari 8. A complete set, up to about three months prior, can be purchased on an external hard drive from the Bureau, though associated documentation notes that:

- No values are reported for the first two hours and last two hours of the day for the period up until 30 June 1994, due to the absence of satellite images at these times during the initial period of operation of GMS4.
- The values are sparser during the period July 2001 to June 2003, which spans the period of reduced imaging frequency at the end of the life of GMS-5, and the initial few weeks of operation of GOES-9 in the Australian region.

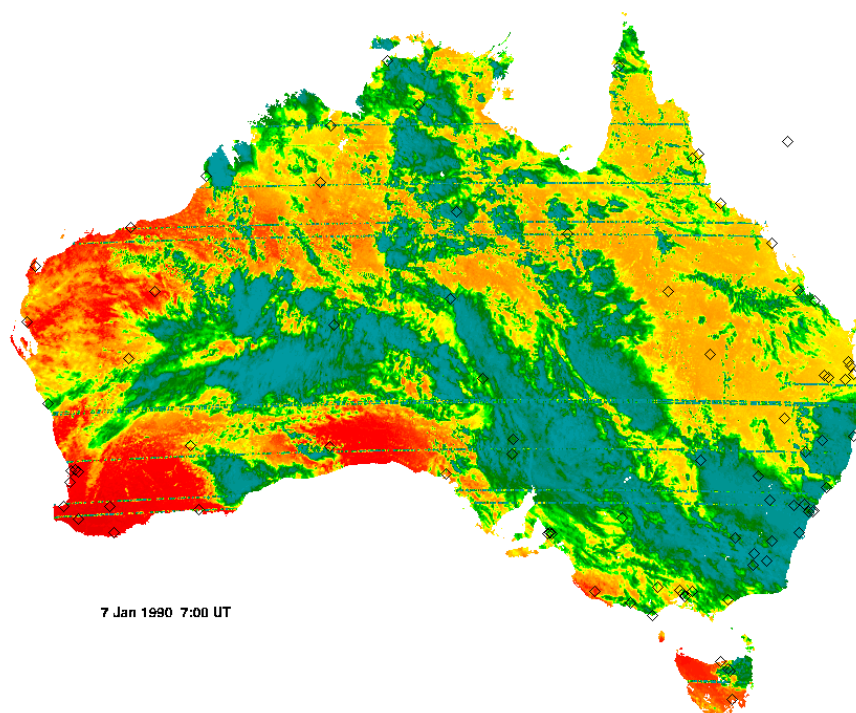
Apart from those restrictions, and occasional gaps, values of both global irradiance ( $G$ ) and direct normal incident radiation ( $R$ ) are given for every hour in which part of the Australian continent is sunlit. The data are provided at  $0.05^\circ \times 0.05^\circ$  resolution, which corresponds approximately to a 5-km grid, and the pixel size in Figure 12. After checking the alignment of the solar radiation images with a detailed outline map of Australia, we identified the five closest pixels to the climate zone reference locations listed in Table 1 and labelled in Figure 1, as shown by diamonds on Figure 12 and Figure 14.

The spatial coverage was also extended to around 50 km offshore, seemingly by extrapolation of irradiance computed over land. For ease of interpretation, only the over-land data are shown in the figures below.



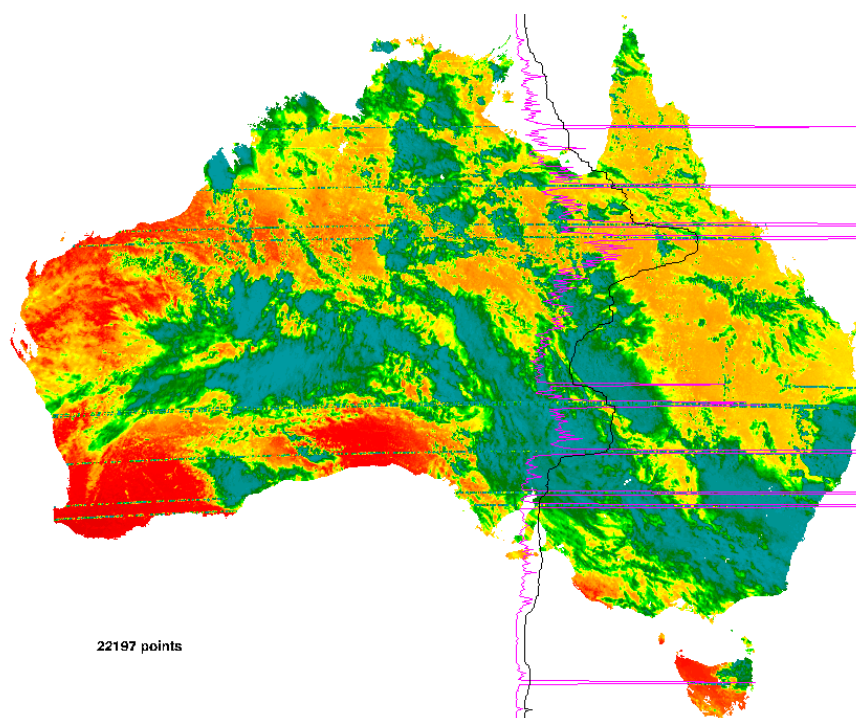
#### 4.4 Spatial interpolation of erroneous data

Extension of the dataset back to 1990 included many satellite images that contain incorrect data. An example is shown in Figure 12, from software used to review the dataset.



**Figure 12.** Satellite-derived direct normal irradiance for 7 January 1990, 07:00 UT.

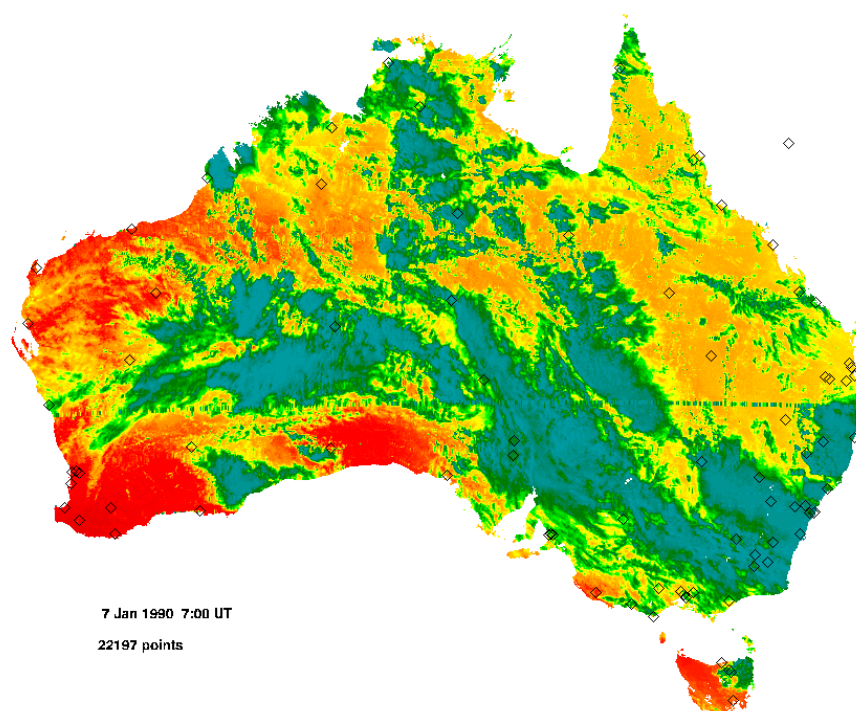
In past work, we developed an algorithm to detect such errors, as shown in Figure 13.



**Figure 13.** Satellite-derived direct normal irradiance as in Figure 12, showing error detection. Lines illustrate the algorithm described herein, and the number of pixels affected is shown.

The lines through the image look like ‘static’ on a television screen, and they arise in a similar way, as disturbance of one or more scan lines in the satellite image. Because of the geometry of the satellite camera and the projection to latitude and longitude, the lines are not straight. We converted back to satellite image coordinates to straighten the lines and simplify error detection.

Figure 13 illustrates results from the algorithm, which first calculates a new image by smoothing in the vertical direction, then measuring the difference between raw and smoothed images. The variance of this difference along the (straightened) horizontal lines gives values illustrated by the purple line. Those values are smoothed, doubled, and offset to give the black line, which serves as a threshold. Where the purple line crosses to the right of the black line, the data along that near-horizontal arc are replaced by the smoothed value, as in Figure 14.



**Figure 14.** Satellite-derived direct normal irradiance as in Figure 12, after correction. Some errors remain, such as around the NSW/Qld border. Diamonds denote NatHERS representative sites.

Most of the seriously anomalous data are detected by this algorithm, but some visible anomalies miss detection or, as in Figure 14 around the Queensland-NSW border at 29° S latitude, are incompletely removed. As in this instance, those that miss detection or complete correction are mostly not very numerically wrong even though the spatial pattern looks distinctive.

## 4.5 Comparison with ground-based data

The satellite images are labelled in UT hours. Conversion of these times to MST as used in NatHERS is inexact for two reasons. One is that the NatHERS values are totals for the hour centred on the specified time, whereas the satellite-derived data are instantaneous measures. The second is that the satellite instruments scan the Australian continent, with different times for each pixel. According to the supplied metadata, the observation time in minutes after the start of the hour varies smoothly with latitude for each satellite and hour of the day, but differs between satellites and, for some satellites, between hours of the day. Times for any latitude are interpolated from Table 3, which gives them at 5-degree latitude increments (Weymouth and Le Marshall 2001, and supplied

metadata). For example, the data in Figure 12 come from GMS-4 A imagery, so the times range from 07:46 to 07:52 UT, North to South. For the earlier file labelled 0400 UT, from GMS-4 B imagery, observation times range from 04:39 to 04:45 UT.

**Table 3. Minute offset within the nominal hour for satellite-derived solar radiation data.** In earlier data, for UT hours labelled B, times are shifted back by 6 – 12 minutes as shown.

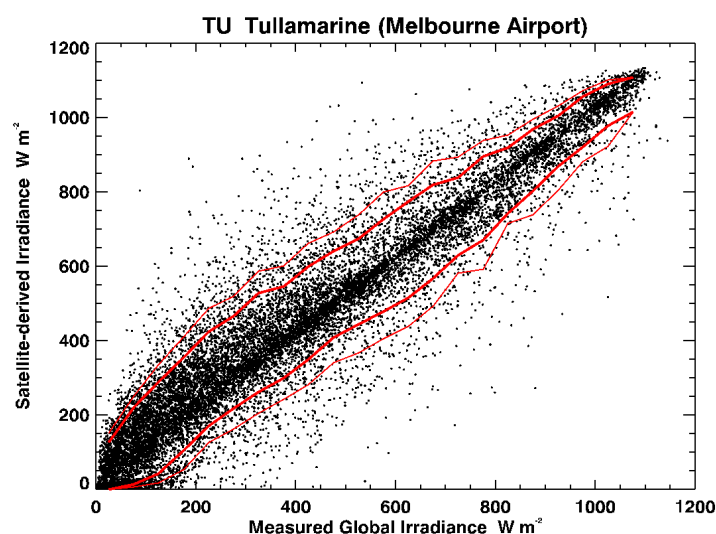
Latitude	GMS-4	GMS-4	GMS-4	GMS-5	GOES-9	MTSAT-1R	MTSAT-2	Himawari-8
Start date	1990/01/01	1993/01/01	1994/07/01	1995/06/11	2003/05/21	2005/11/01	2010/07/01	2016/03/22
End date	1992/12/31	1994/06/30	1995/06/10	2003/05/20	2005/10/31	2010/06/30	2016/03/21	Ongoing
-10.0	45.7	47.2	46.7	46.7	39.9	46.2	44.7	36.0
-15.0	47.7	48.2	47.7	47.7	41.0	47.2	45.7	36.9
-20.0	47.7	49.3	48.8	48.8	42.0	48.3	46.8	37.0
-25.0	48.7	50.2	49.7	49.7	43.0	49.2	47.7	37.9
-30.0	49.6	51.1	50.6	50.6	43.9	50.1	48.6	38.4
-35.0	50.5	52.0	51.5	51.5	44.7	51.0	49.5	38.6
-40.0	51.2	52.7	52.2	52.2	45.5	51.7	50.2	38.9
-44.0	51.8	53.3	52.8	52.8	46.0	52.3	50.8	39.1
B shift	-7.0	-6.5	-6.2	-7.0	-12.0	0.0	0.0	0.0
A: UT hours	18 19 20 21	23 00 01 02 03	05 06 07 08 09	11				
B: UT hours		22	04	10				

According to metadata supplied by the Bureau with the initial release, the data have been checked against the 1-minute measurements from Bureau instruments described in Section 4.1 above:

- The mean bias difference (average of the satellite - surface difference), calculated on an annual basis across all surface sites, is +11 to +40 W m<sup>-2</sup> and typically around +20 W m<sup>-2</sup>. This is +4% of the mean irradiance of around 480 W m<sup>-2</sup>. The root mean square difference, calculated on a similar basis, is around 130 W m<sup>-2</sup>, which is 27% of the mean irradiance.
- It should be noted that a particular [satellite-derived] value may not be representative of a 1-hour period, due to variations in the solar zenith angle during the hour, and most significantly because of variations in atmospheric conditions such as cloudiness.

To confirm the suitability of the satellite-derived estimates, we looked at their correlation with ground-based measurements in the NatHERS files (MST), as shown in Figure 15, reproduced from the 2016 report (Liley, J Ben 2017) and also the 2005:2016 comparison (Liley, Ben 2019). The correlation is closest when the satellite-derived values are interpolated from measurement minute to the centre of the hour in the ground-based data; it was then comparable to the figures quoted above.

Figure 15, as Figure 18 in Liley (2017), prompted some concern that, though the averages agree, actual values are within 20% only 90% of the time, and they sometimes disagree by ~40%, or over 200 W m<sup>-2</sup>. In explanation, it should be noted that the satellite-derived data are for cells of size 0.05° x 0.05° in latitude and longitude, or approximately 5 km x 5 km. Agreement is measurably better for the correct grid cell than for any adjacent ones, and differences are readily attributable to spatial averaging of satellite data and temporal averaging of ground-based.

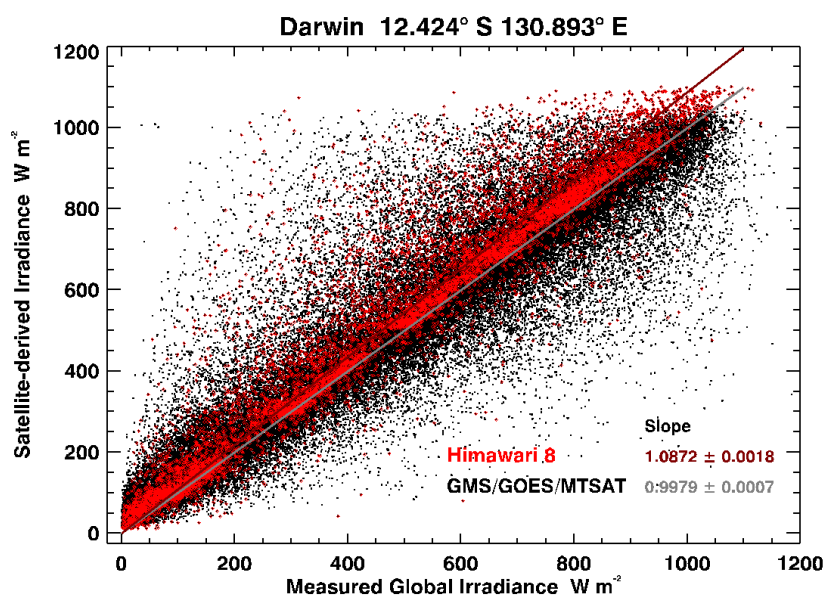


**Figure 15.** Measured hourly and satellite-derived instantaneous global irradiance interpolated to the same times at Tullamarine. Red contours are 5<sup>th</sup>, 10<sup>th</sup>, 90<sup>th</sup>, and 95<sup>th</sup> percentiles.

This question was fully explored in Liley (2019), which describes the creation of a pair of RMYs for Tullamarine, one with ground-based and the other with satellite-derived solar data. As summarised in that report, simulations by CSIRO of the Software Accreditation Protocol benchmark house design 101, modelled in North and South orientations using the two ‘test’ RMYs, found negligible differences in heating and cooling loads, and no difference in Star rating to the first decimal place.

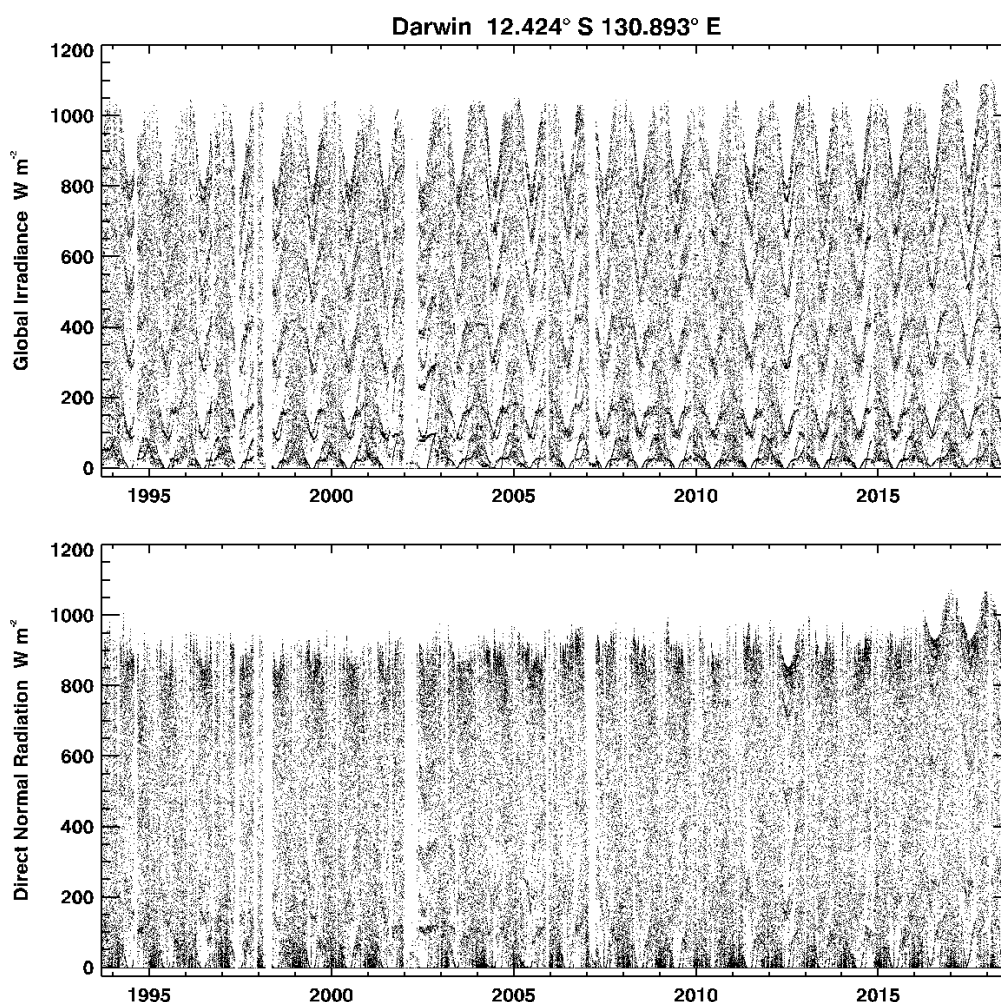
#### 4.6 Himawari 8 data

As Table 3 shows, all satellite data from 22 March 2016 are from Himawari 8, the latest in the GMS series. The new satellite provides imagery every 10 minutes, but the Bureau radiation data time series continue as hourly, using the Himawari 8 scans nearest in time to the prior satellite’s times.



**Figure 16.** Measured hourly and satellite-derived instantaneous global irradiance interpolated to the same times at Darwin. Black symbols, grey fitted line, and fitted slope are for earlier data, while red symbols, line, and slope refer to Himawari 8 data.

When post-2015 satellite-derived data, mostly from Himawari 8, were appended to the earlier time series, recent data were found to be markedly higher at some sites. The problem is illustrated in Figure 16, which compares satellite-derived data before (black) and after (red) 22 March 2016 with simultaneous ground-based measurements at Darwin. The difference is even more apparent in time series plots, as shown in Figure 17.



**Figure 17.** Time series of global and direct normal-incident radiation at Darwin from satellite-derivation. Data after 22 March 2016 are from Himawari 8.

The problem is not uniform across the continent, with most sites showing good agreement. Whereas the fitted slope of  $1.0872 \pm 0.0018$  in Figure 16 for Darwin in the Himawari 8 era is nearly 9% high, the same analysis for Tullamarine (not shown) gives a slope of  $0.9723 \pm 0.0018$ . That is very close to (and slightly less than) the slope of  $0.9832 \pm 0.0007$  for the earlier data at Tullamarine. At all other radiation sites, the mean ratio of Himawari 8 data to ground-based is somewhat greater than for previous data, but the Himawari 8 data may be valid; at many sites, the fitted slopes are closer to unity for Himawari 8 data. Nonetheless, as illustrated by Figure 17, the difference is greater for direct normal incident radiation, and the step in time series is itself a problem, as it was for wind.

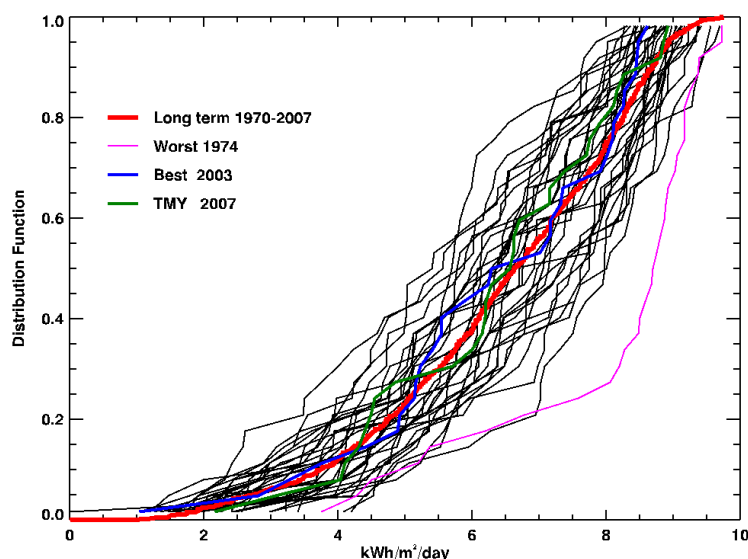
This issue was unfamiliar to the Bureau, and they had no ready explanation. Because of this uncertainty, and extensive testing already undertaken with the 2016 datasets, the NatHERS Steering Committee decided to postpone adoption of the 2019 update, and use the 2016 files in the National Construction Code (NCC) 2022 update cycle.



## 5 Derivation of RMY/TMY files

### 5.1 Finkelstein-Schafer statistics

The construction of RMYs from the NatHERS data follows the prescription of Marion and Urban (1995) for Typical Meteorological Years (TMYs), with some refinement as described in Liley et al. (2008). Specifically, the selection depends on Finkelstein-Schafer (F-S) statistics, which can be understood from Figure 18.



**Figure 18.** Distribution functions of January daily global irradiance for Auckland, New Zealand. The best match to long-term distribution for irradiance alone is 2003, whereas the TMY year for January (2007) was chosen from a weighted mean, as in Table 4.

For each month, the distribution of values for a variable in that month of each year is compared with the overall distribution for that month in all years. The F-S statistic measures total absolute differences in the vertical direction, corresponding to probability rather than physical values, so F-S values of different physical quantities can be compared or combined. The more familiar concept of measuring departure from some average along the horizontal axis of Figure 18 (i.e., in  $\text{kWh m}^{-2} \text{day}^{-1}$ ) would require normalisation by standard deviation, interquartile range, or similar measure.

**Table 4.** Weightings for Finkelstein-Schafer statistics in NatHERS RMYs.

Index	RMY
Max Dry Bulb Temperature	1
Min Dry Bulb Temperature	1
Mean Dry Bulb Temperature	2
Max Dew Point Temperature	1
Min Dew Point Temperature	1
Mean Dew Point Temperature	2
Max Wind Speed	1
Mean Wind Speed	1
Global Radiation	5
Direct Radiation	5
Total (denominator)	20

By contrast, the F-S statistics can be combined directly, with any preferred weightings. For the RMYs (and TMYs) created for NatHERS since 2012, the weightings used are as shown in Table 4.

## 5.2 Ambiguity in the Sandia method

To this point the prescription is unambiguous, but the next step is not. The five months with lowest combined F-S score are to be ranked in order of “closeness of the month to the long-term mean and median” (Marion and Urban 1995). Marion and Urban do not say how they compare these two measures, nor how they weight them for the different parameters as both mean and median are expressed in physical units so require some normalisation. There does not seem to be any standard for resolving this question, so we employ a method previously developed by NIWA.

For the New Zealand TMYs (Liley, J B, Sturman et al. 2008), after exploring several techniques, we developed a modified ‘signed’ F-S statistic that gives the desired central tendency, and can be combined in a weighted sum exactly as can the F-S statistic. Thus, we use the standard F-S weighted sum to obtain the best five months, and then the modified statistic to rank them, subject to completeness of data not already included in the weighting. The further step to limit the number or length of ‘runs’ is handled in a related manner. Mathematical details are given in Appendix B.

## 5.3 EnergyPlus TMYs

In the 2012 revision of NatHERS, the RMYs in the amended ACDB format, also called ‘AccuRate’, for the NatHERS software that uses it, were converted to TMYs in the EnergyPlus format (\*.EPW files) preferred by many building energy researchers. That conversion was imperfect, in that AccuRate follows a convention for the effective time zone (of radiation data) based on historical use of Mean Solar Time by the Bureau (Walsh, Munro et al. 1983). By contrast, EnergyPlus radiation data should be centred on the hour preceding the time stamp of the other climate variables. In the present work, the 2016 files were generated for EnergyPlus with the correct time alignment, but for the same representative years. In consequence, the other climate variables will be identical between the RMYs and TMYs, but the radiation data will differ slightly, appearing as offset by some minutes. Because of the way that the ground-based radiation data are aggregated, and the satellite-derived data are interpolated, daily totals will be the same for both datasets, within rounding errors.

EnergyPlus files also contain provision for many other climate variables. As a component of this revision, we have added the fields that provide information on precipitation, using rainfall data from the climate stations.

Both the 1990-2015 and 1990-2018 time series can be provided, but the latter carry the same confusion about how to interpret the change to Himawari 8. For provision to researchers, we anticipate that the TMYs corresponding to NatHERS 2016 will be used.

From past agreement with the NatHERS team, the 2012 TMY (.EPW) files are supplied from a NIWA server to registered users. The list of registrants contains more than 250 email addresses of those who are likely to wish to know of the revised files. Should the NatHERS so wish, the 2016 TMY files could be served in a similar way.

## 6 Review with Bureau of Meteorology

On Tuesday 12 November 2019, Ben Liley visited the Bureau of Meteorology in Melbourne to review the data choices, quality control algorithms, and subsequent analyses in developing climate time series for creating RMYs and TMYs. The discussion was organised by Dr Brad Murphy, Head of Climate Data and Analysis within the Bureau's National Forecast Services, who had previously answered queries related to climate stations and datasets for NatHERS. Brad included members of the satellite applications team, climate data experts, the head of the Bureau's data QC unit, and members of the radiation group who have collaborated with NIWA's Lauder group for over 30 years.

The purpose of the meeting was to confirm that relevant Bureau staff endorsed the data sources chosen and the methods used by NIWA, and that relevant staff understood enough to potentially reproduce that work in future. For NIWA, Ben Liley has led work in this area for NatHERS for the past decade, but dependence on one person constitutes a weakness for a product on which the work of many others depends. The Bureau supply all the raw data, and they have procedures for quality control and conversion of physical parameters. It could improve the security of future NatHERS climate data work if the Bureau could supply climate time series ready for deriving RMYs, TMYs, or other datasets that NatHERS may require. Against that, the Bureau would need specific funding to dedicate any of their staff to familiarise themselves with the tools NIWA has developed, or to reproduce the results with their own tools.

With those considerations, Ben initially gave a presentation to the group, describing and elaborating on all aspects of this report and preceding work for NatHERS. That was followed by meetings with individuals or small groups to discuss different aspects of the work.

Bureau staff expressed no reservations about any of the procedures applied; indeed, they saw potential value to the Bureau in adopting some of the techniques. Certainly it would be most cost-effective for NatHERS if the Bureau were to automate such error detection in its data handling, so that it could supply quality-controlled data to the required standards at little or no extra cost. At present, the Bureau supplies data of most types (temperature, pressure, dew point, wind speed and direction) as stored in their climate databases after limited quality control. These are such things as absolute limits on range, or data flags already set by human scrutineers. A difficulty for routinely applying tests like those used for the NatHERS data is that, for the most part, we are more concerned with typicality than with preserving any extremes. It would not be a problem for NatHERS if the odd genuinely extreme value were classified as erroneous, but it certainly would be unacceptable for Bureau records.

Thus, a more likely scenario would be that the Bureau would, on request from NatHERS, extract the required datasets and apply the filters and masking that we have found necessary. As the Bureau would implement that within their own data processing systems, it would require some time and effort to recode the detailed NIWA algorithms for that purpose. Alternatively, NIWA could be contracted to provide such code in any required language, but that is not a small task.

The present code is written in IDL (<https://www.harrisgeospatial.com/Software-Technology/IDL>), originally 'Interactive Data Language', which is widely used in geophysical sciences and image processing, including satellite applications. The code is efficient and functional, and the routines could be provided in a form that others can use, but the NatHERS work to date has involved repeated cycles of algorithm development and visualisation of the results, for which IDL is well-suited. However the quality control of data is implemented, those visual checks are essential.



Another possibility, suggested by Bureau staff and worthy of consideration for evolution of NatHERS, would be to use BARRA (Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia, Jakob, Su et al. 2017). A ‘reanalysis’ uses data assimilation to optimally combine observations and model forecasts and so provide the best representation of the atmosphere. Embedded within the global atmospheric reanalysis ERA-Interim, which has a nominal resolution of 80 km, BARRA-R represents the atmosphere at approximately 12 km horizontally. Within several more densely populated domains, BARRA uses model physics to generate values on 1.5 km horizontal grids that satisfy dynamical equations of the atmosphere and honour the land surface characteristics and heterogeneity. This sort of product, or even just the 12 km grid of BARRA-R, could be used to generate time series for the NatHERS reference sites.

It would also provide an ideal route to a possible future version of NatHERS that could use a representative local climate dataset for anywhere in Australia. That would bypass the approximation implicit with climate zones, and their dependence on a tessellation of the continent as provided by post code boundaries. As NIWA has a closely related collaborative development (the New Zealand Convective Scale Model), we would expect to be able to offer relevant expertise if NatHERS does consider this approach.

In discussion with the solar radiation group, Ben demonstrated the erroneous periods in ground-based radiation measurements. Unlike the other ground-based data, to which only basic checks are applied, the radiation measurements are in general closely scrutinised. The list of erroneous periods was gratefully received, and they will hereafter be flagged in the datasets. Bruce Forgan and Lance Passamani also volunteered that there are extensive metadata for the radiation instruments, recording instrument changes, calibration, alignment, cleaning, and noteworthy events. That information will be sought for any future NatHERS updates that we undertake. As noted above, it would also be much preferred to have the actual, measured global irradiance at all times for quality control. For those data – the great majority – when the three radiation measurements pass all checks, the ‘component sum’ global irradiance is easily calculated to replace the measured value.

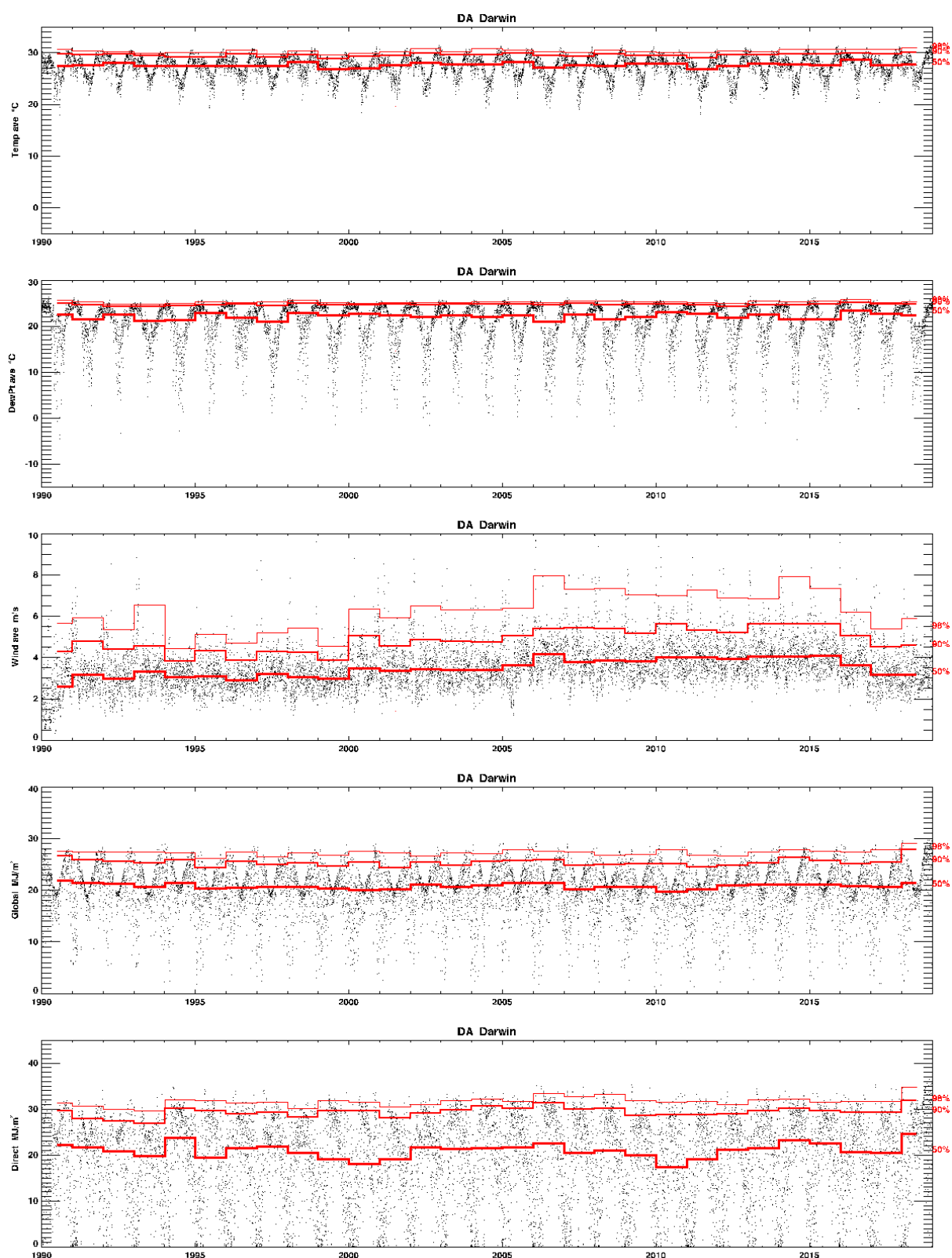
The change in gridded global and direct radiation with the move to Himawari 8 was new to Dr Ian Grant, the Bureau scientist who has led the development of that work for two decades. In particular, Ian had not explored the data as time series at individual locations, so he was very interested to see the code and tools developed by NIWA. As Ian was also familiar with IDL, Ben left him with the transposed data, and the IDL code to extract time series, as well as the code used to implement the corrections described in Section 4.4 above.

Although Ian was keen to explore newfound insights into his *magnum opus*, he knew his time was limited. Over the last nine months, he had been afflicted with pancreatic cancer. Very sadly, on 30 November 2019, Ian lost that battle. He is greatly missed.

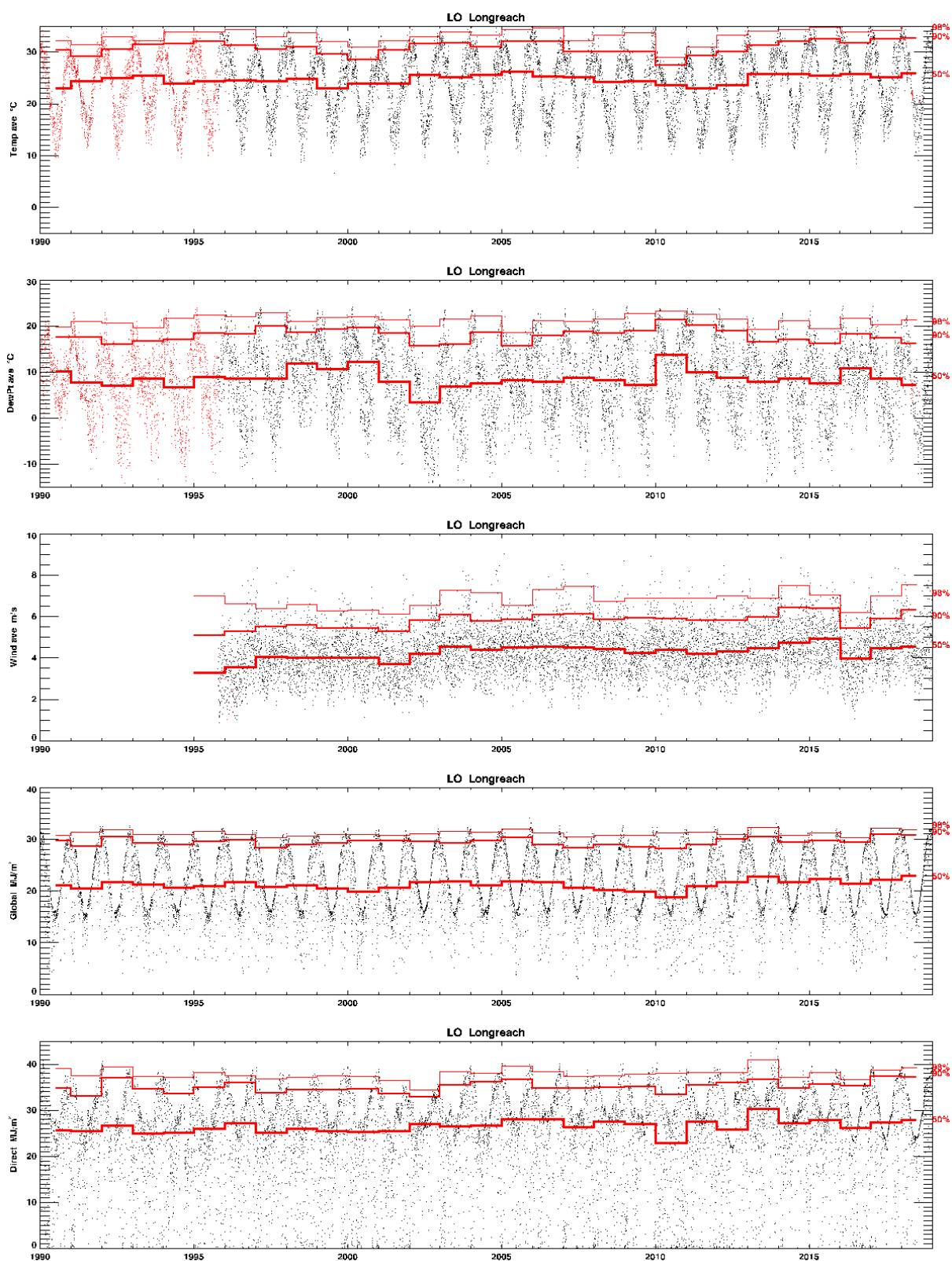
## 7 References

- Chen, D. (2016) AccuRate and the Chenath engine for residential house energy rating. *CSIRO Land and Water*. <https://hstar.com.au/Home/Chenath>
- Delsante, A. (2005) Description of weather data files used by AccuRate. *CSIRO Manufacturing & Infrastructure Technology*.
- Energy Partners (2008) Australian Climate Data Bank - Weather data enhancement for Reference Meteorological Years. *Report to the Australian Department of Environment, Water, Heritage and the Arts, Canberra*: 86.
- Grant, I.F. (2009) Near-real time satellite products to drive Australia-wide land surface monitoring and modelling of surface water and energy balance. In: S. Jones & K. Reinke (Eds). *Innovations in Remote Sensing and Photogrammetry*. Springer, Berlin: 161-172. 10.1007/978-3-540-93962-7\_13
- Jakob, D., Su, C.-H., Eizenberg, N., Kociuba, G., Steinle, P., Fox-Hughes, P., Bettio, L. (2017) An atmospheric high-resolution regional reanalysis for Australia. *The Bulletin of the Australian Meteorological and Oceanographic Society*, 30: 16-23.
- Liley, B. (2019) Climate analysis of 2005 and 2016 NatHERS files. *NIWA Client Report*, 2019176WN: 27.
- Liley, J.B. (2013) Australian climate data and Reference Meteorological Years for NatHERS 2012. *NIWA Client Report*, LAU2013-02-JBL: 35.
- Liley, J.B. (2017) Creation of NatHERS 2016 Reference Meteorological Years including Maleny and Christmas Island. *NIWA Client Report*, 2017103WN: 67.
- Liley, J.B., Sturman, J., Shiona, H., Wratt, D.S. (2008) Typical Meteorological Years for the New Zealand Home Energy Rating Scheme. *NIWA Client Report*, LAU2008-01-JBL: 46.
- Marion, W., Urban, K. (1995) User's Manual for TMY2s (Typical Meteorological Years), NREL/SP-463-7668: 55. <http://rredc.nrel.gov/solar/pubs/tmy2/>
- Walsh, P.J., Munro, M.C., Spencer, J.W. (1983) An Australian climatic data bank for use in the estimation of building energy use: 12. [http://users.tpg.com.au/t\\_design/Trnaus/walsh.PDF](http://users.tpg.com.au/t_design/Trnaus/walsh.PDF)
- Weymouth, G.T., Le Marshall, J.F. (2001) Estimate of daily surface solar exposure using GMS-5 stretched-VISSR observations. The system and basic results. *Australian Meteorological Magazine*, 50: 263-278.

## Appendix A Example Time Series

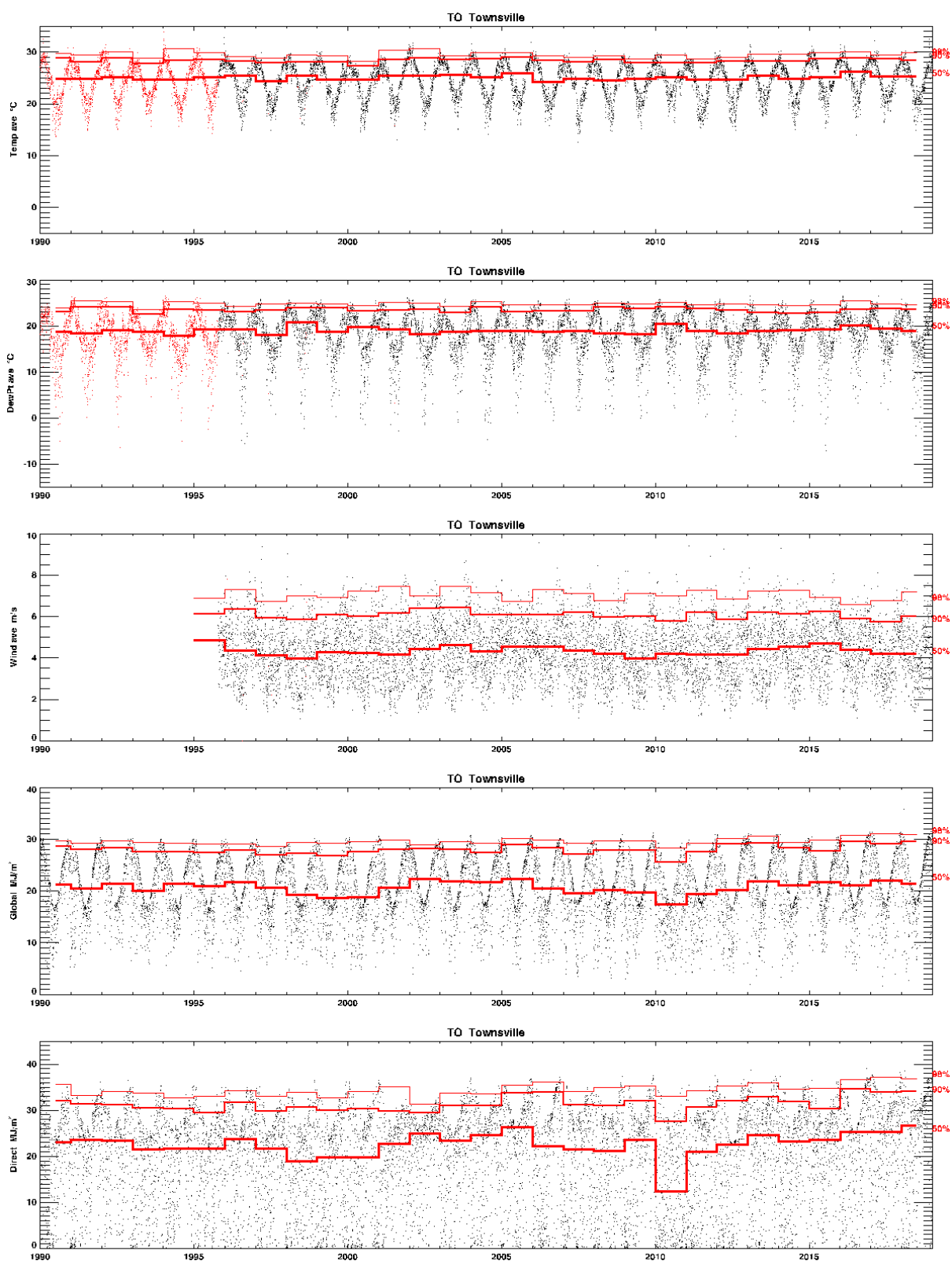


Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#01 Darwin.



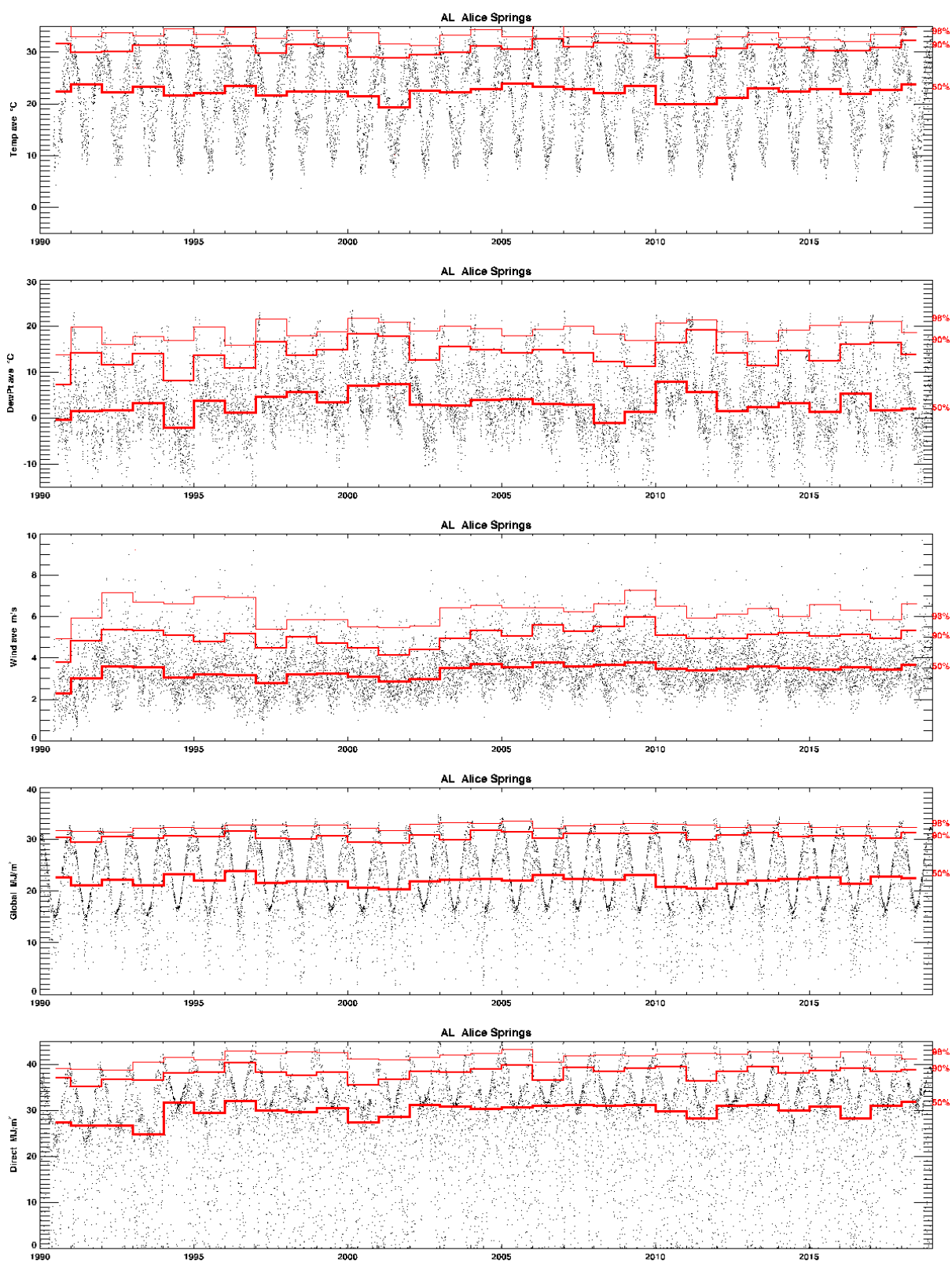
**Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#03 Longreach.**

Red points, spatially interpolated from nearby sites, are used only if required.

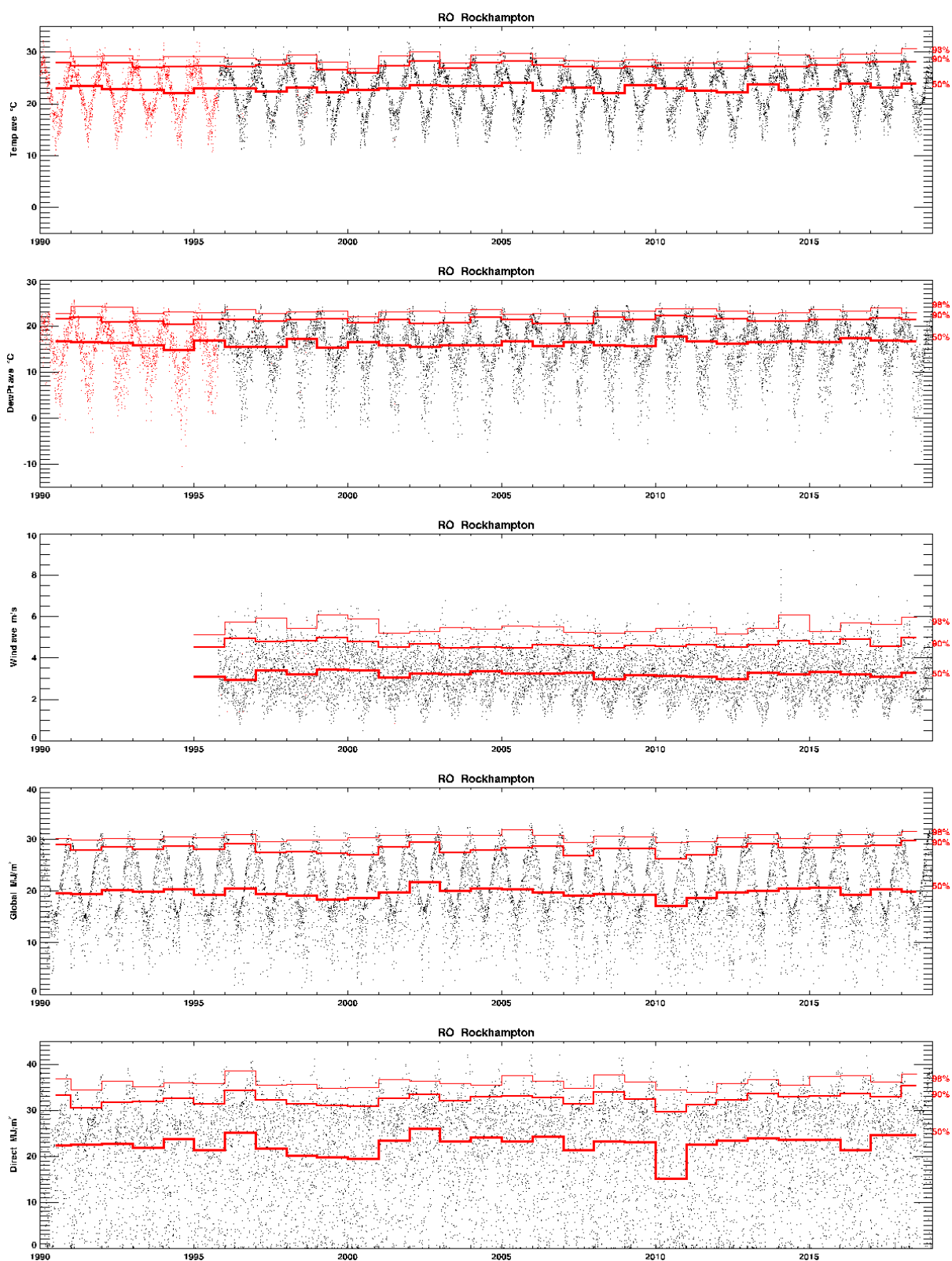


**Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#05 Townsville.**

Red points, spatially interpolated from nearby sites, are used only if required.

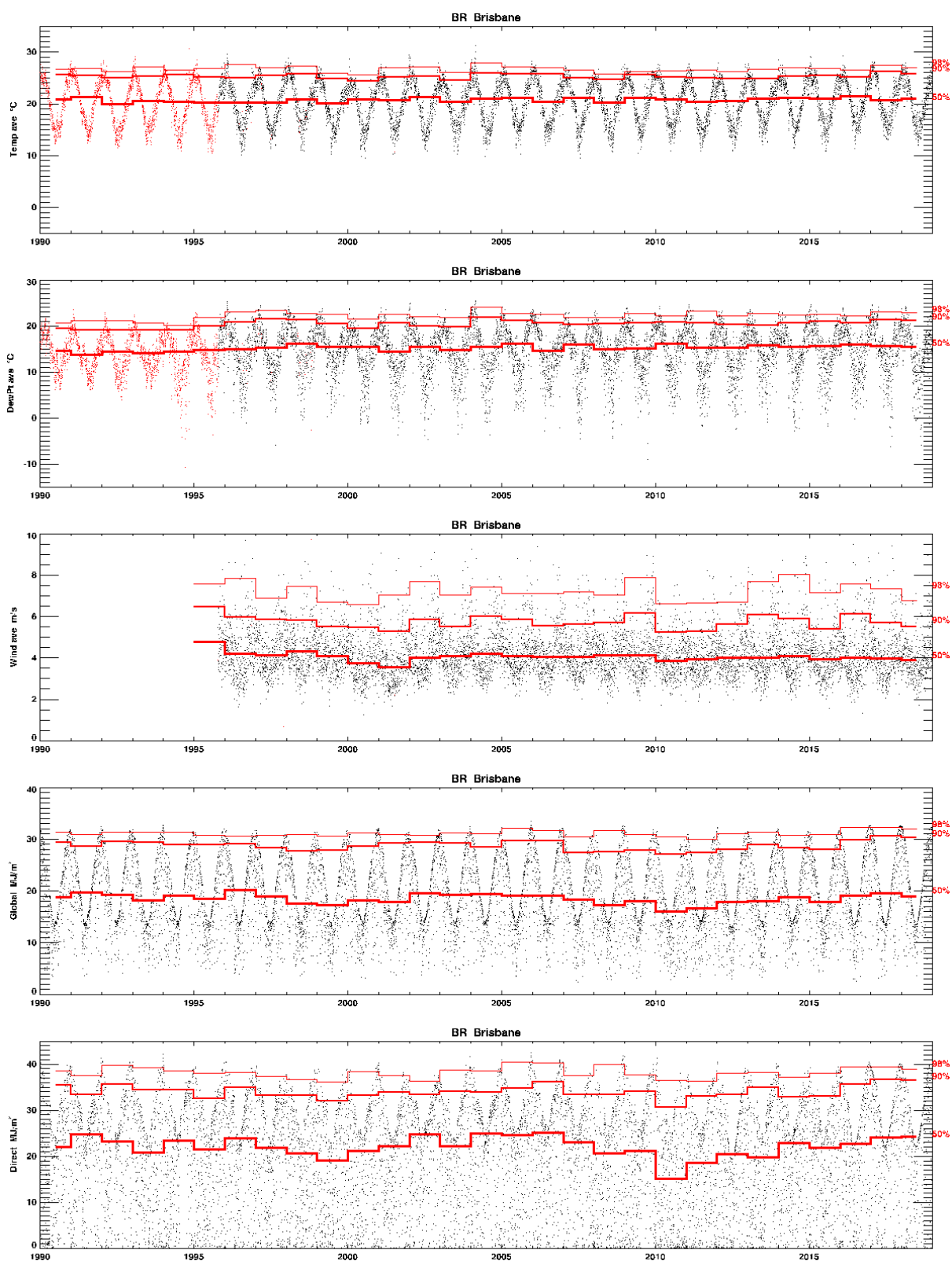


Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#06 Alice Springs.



**Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#07 Rockhampton.**  
 Red points, spatially interpolated from nearby sites, are used only if required.

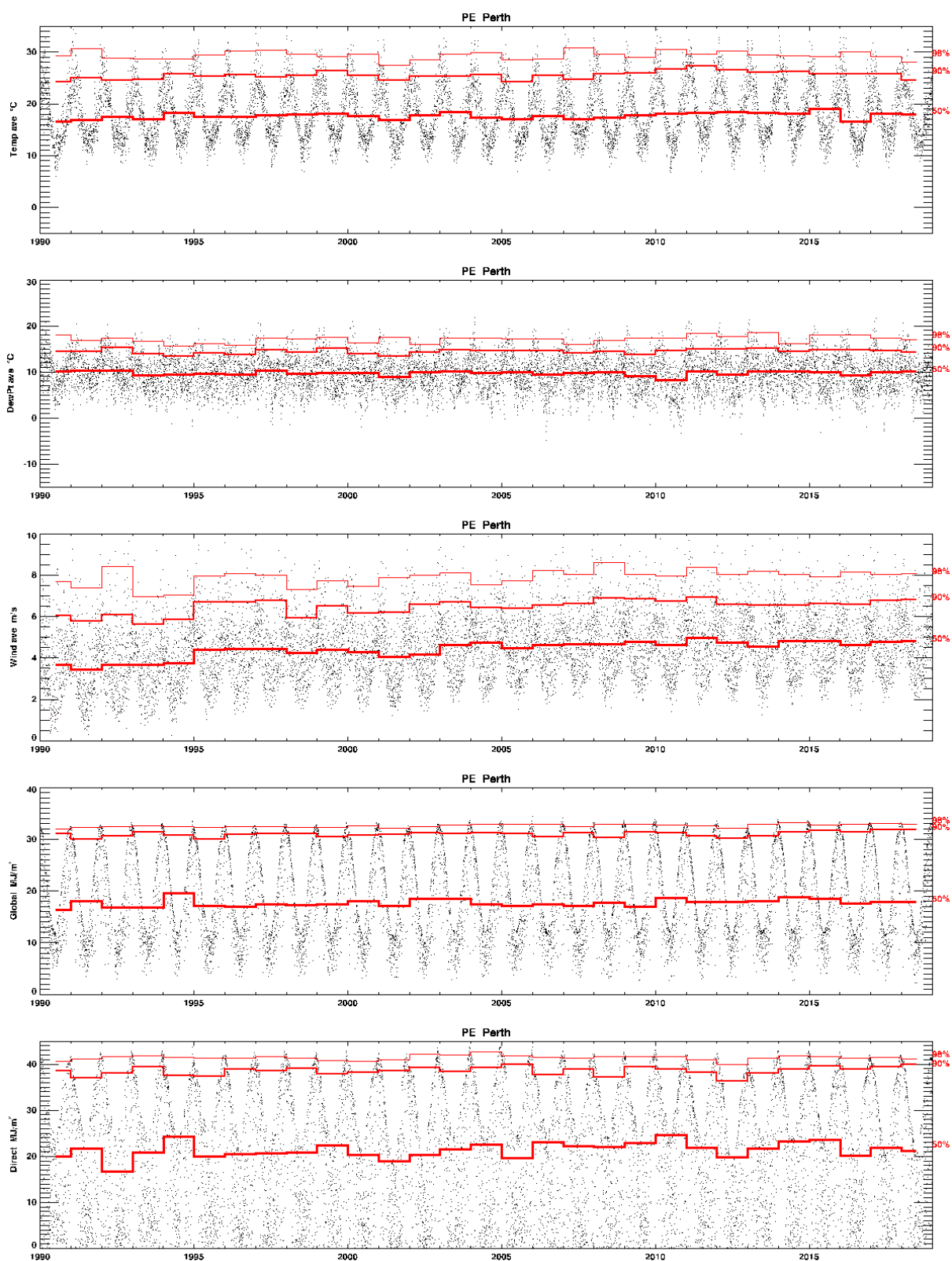




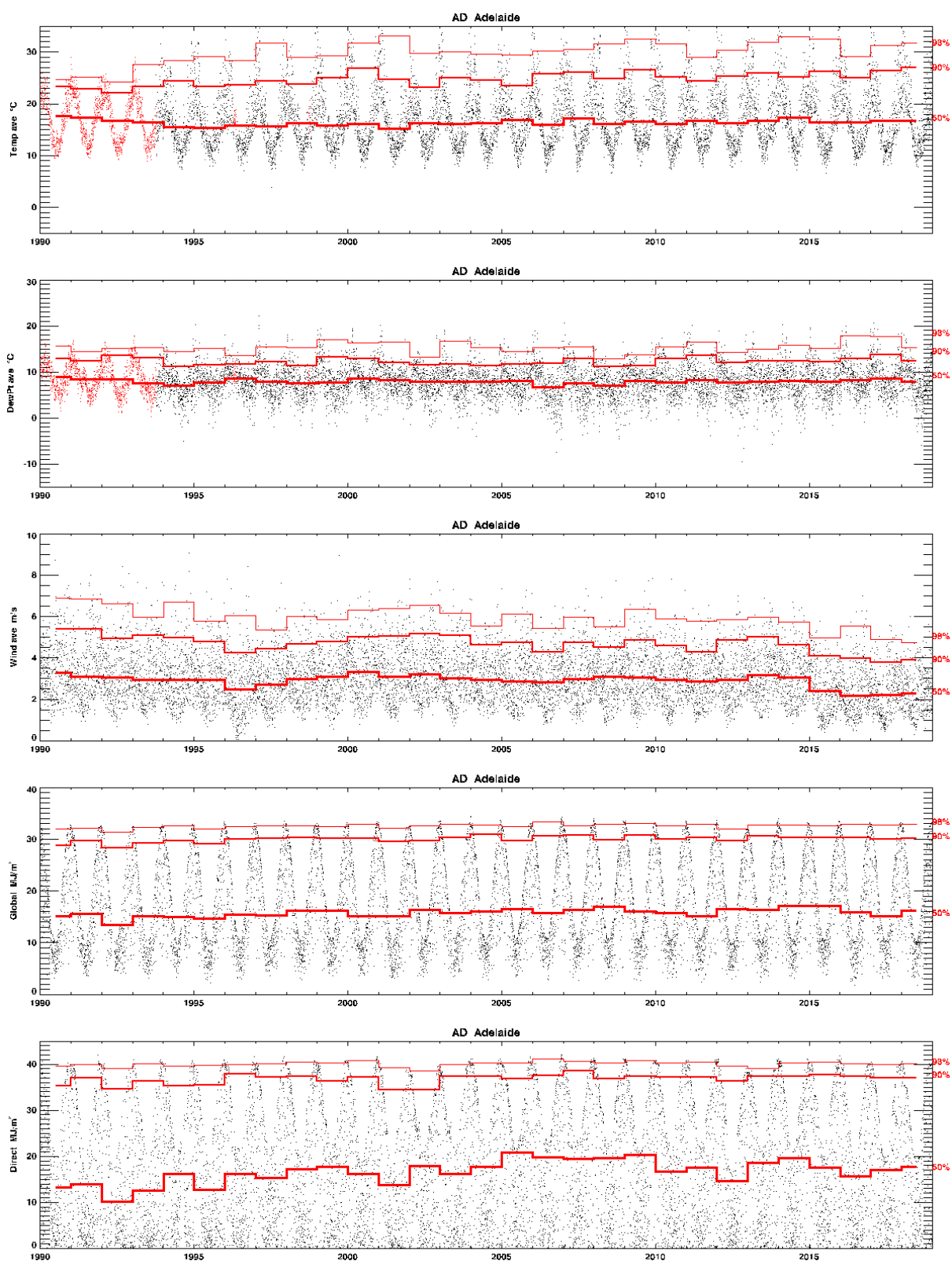
**Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#10 Brisbane.**

Red points, spatially interpolated from nearby sites, are used only if required.



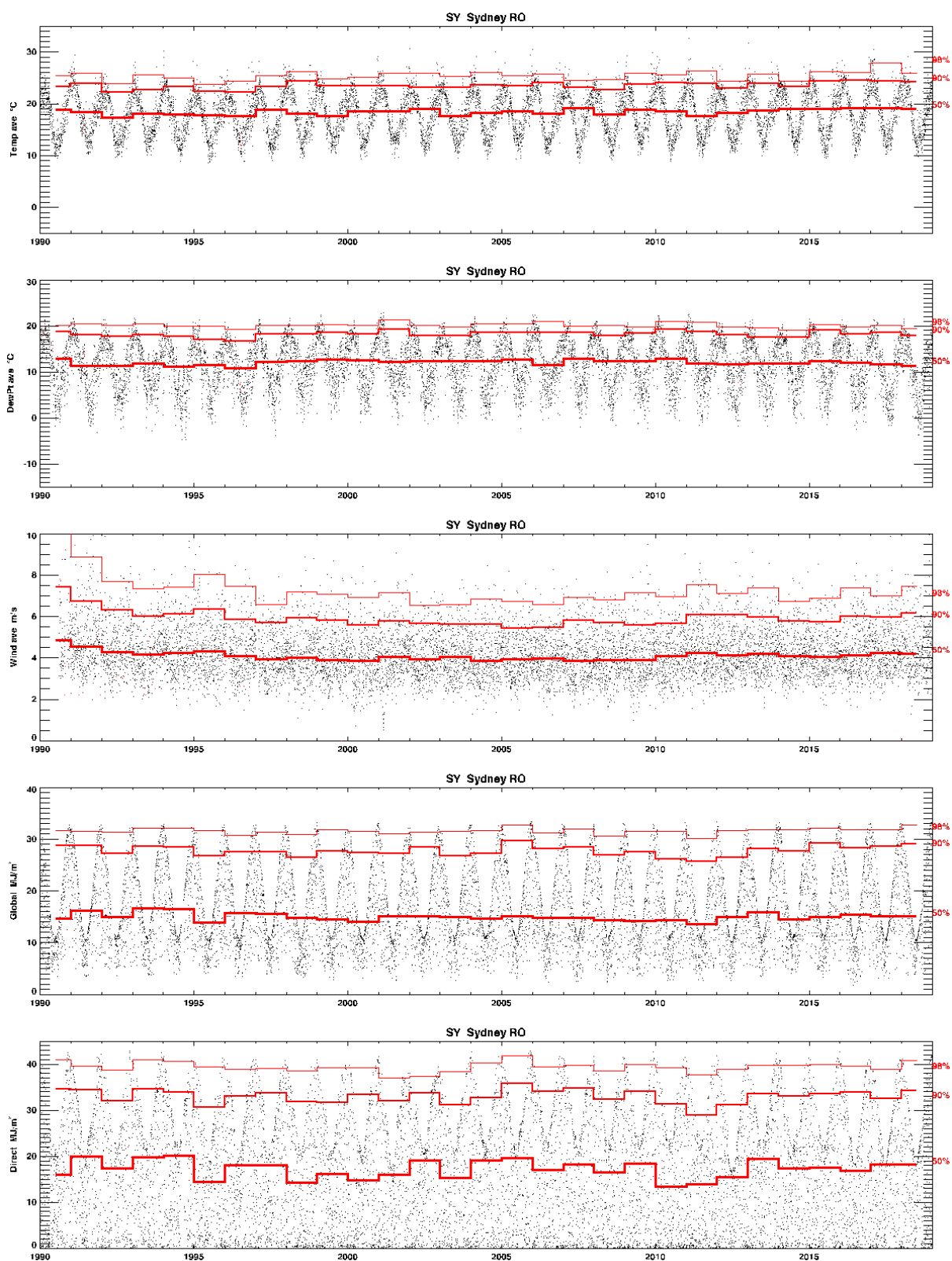


Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#13 Perth.

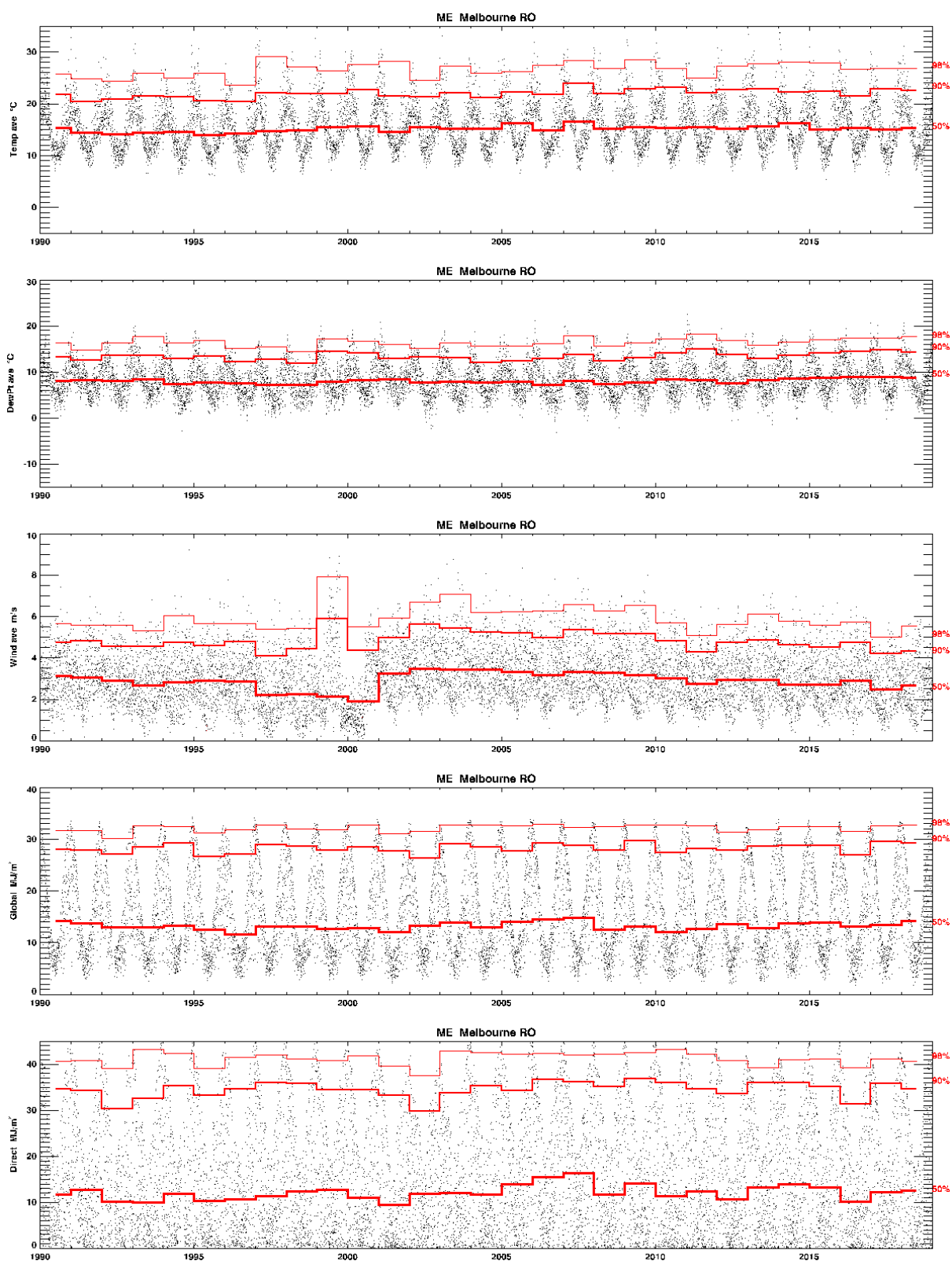


**Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#16 Adelaide.**

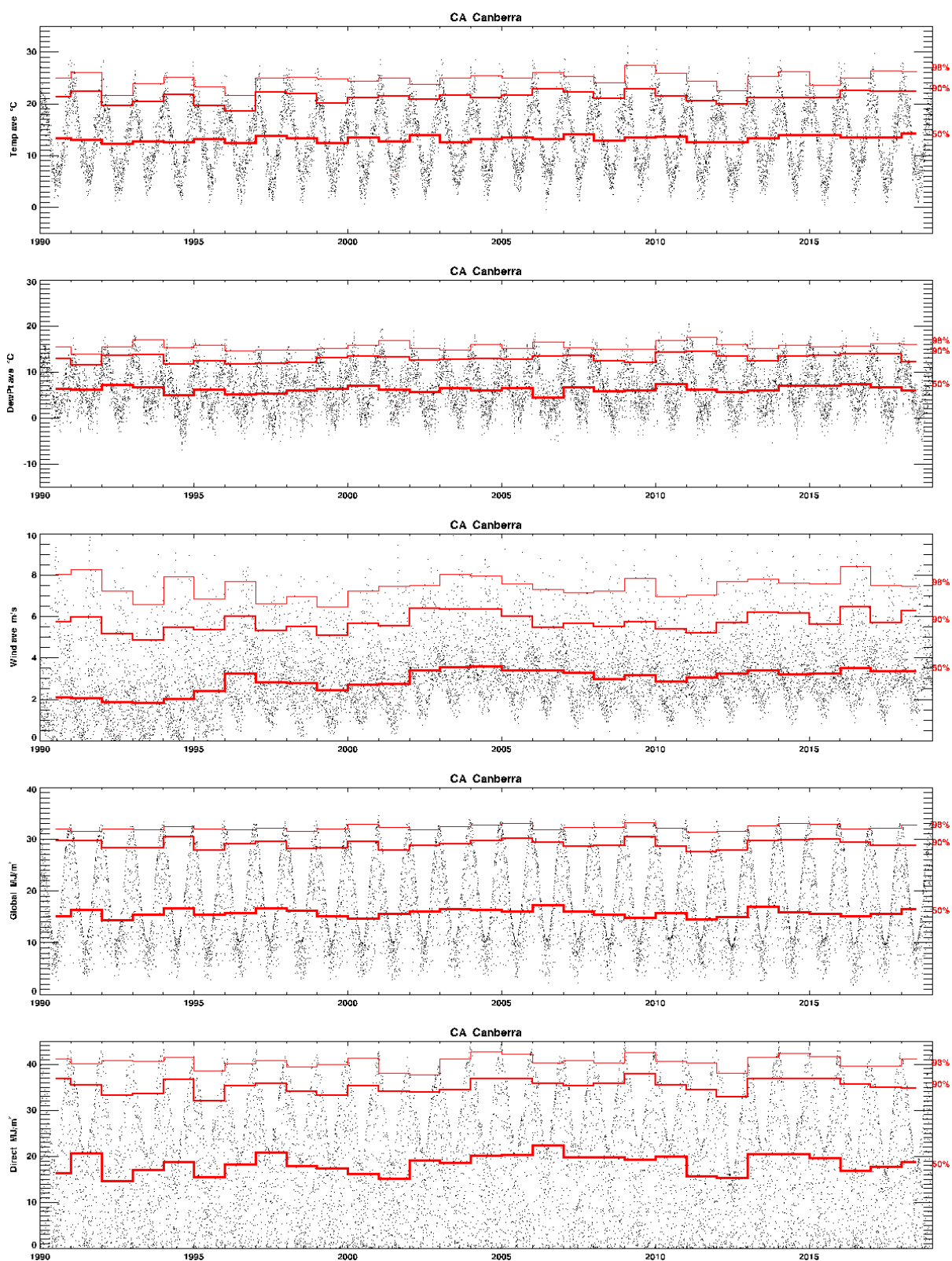
Red points, spatially interpolated from nearby sites, are used only if required.



Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#17 Sydney.

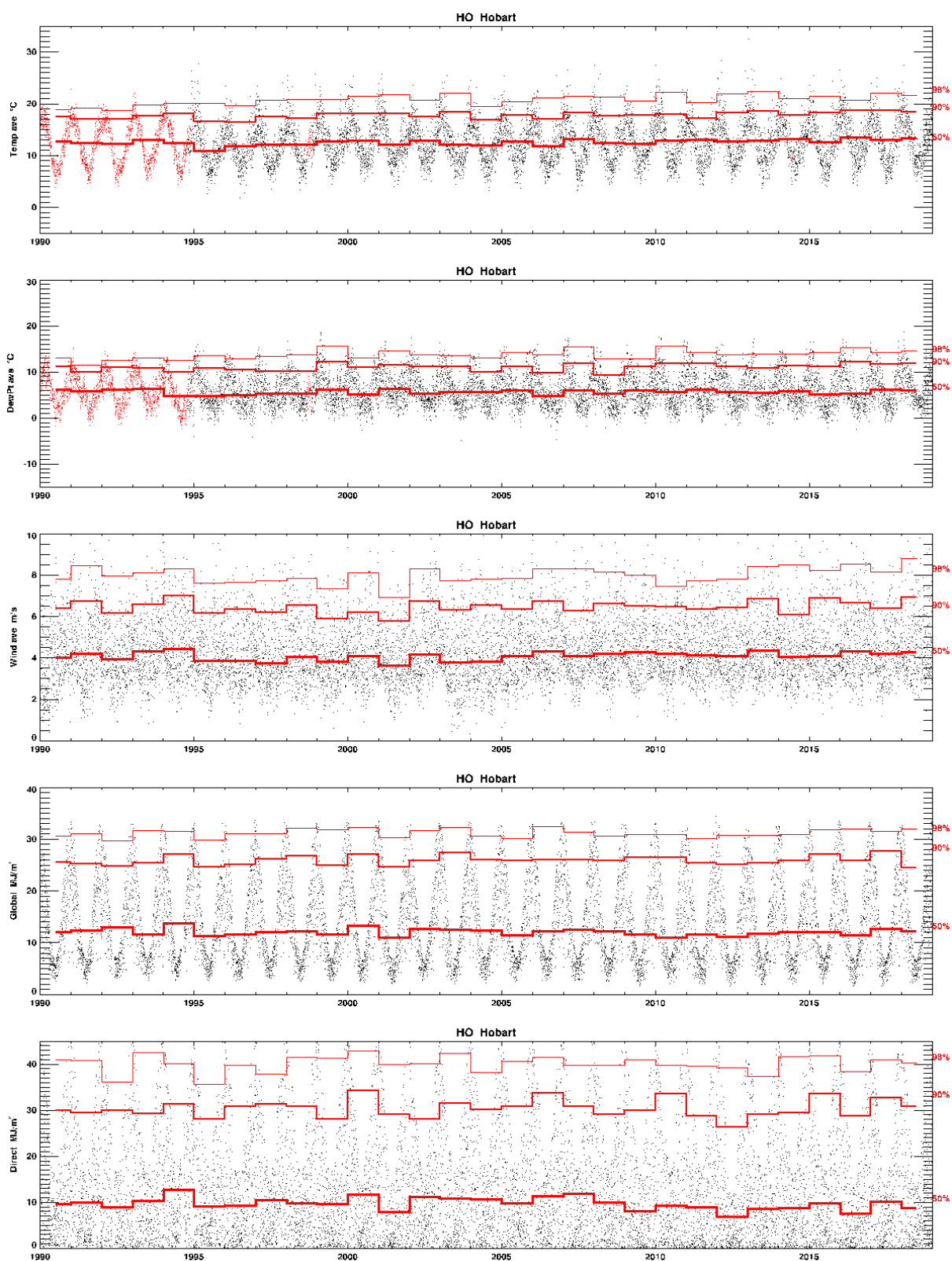


Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#21 Melbourne.



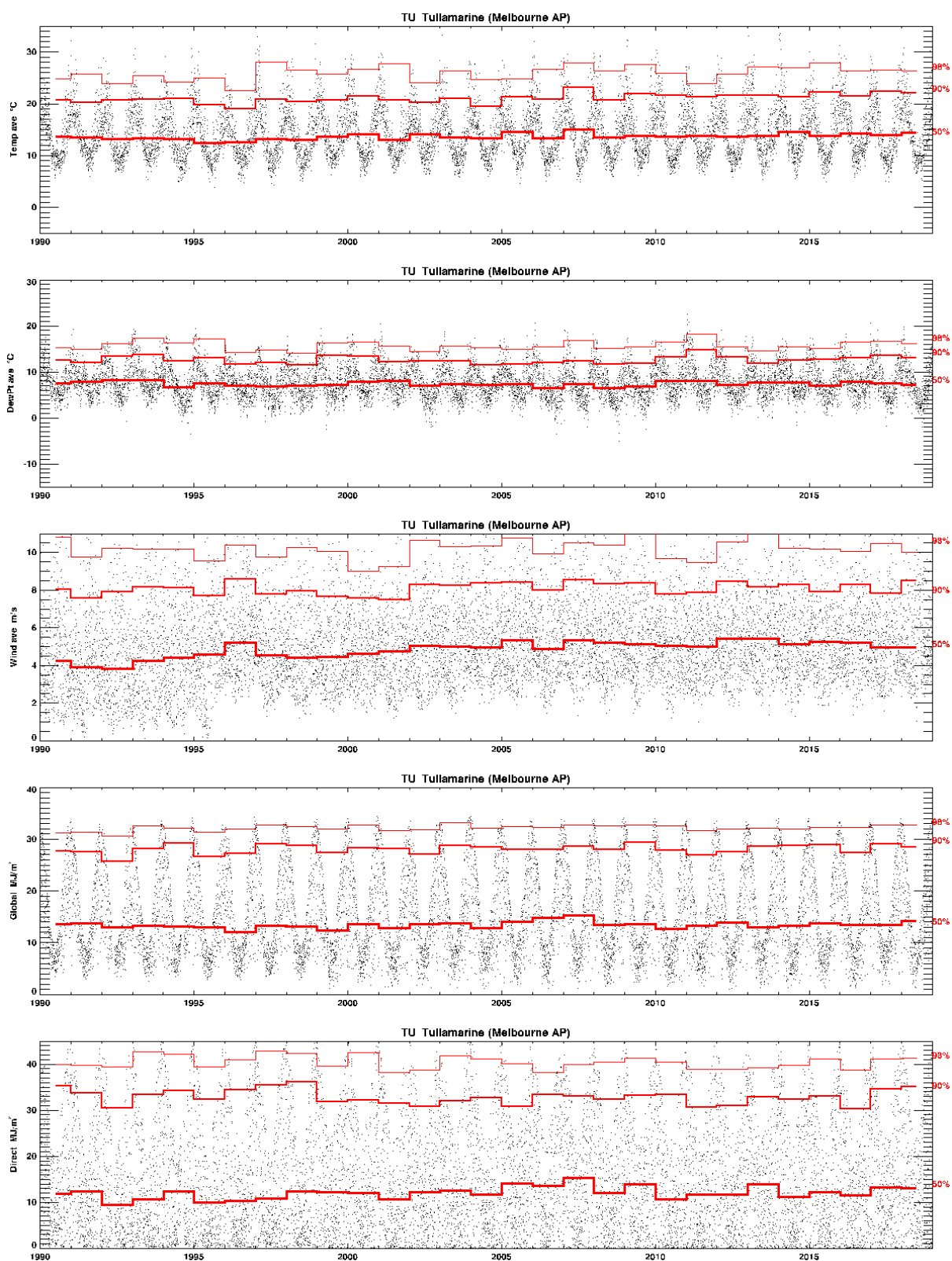
Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#24 Canberra.





**Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#26 Hobart.**

Red points, spatially interpolated from nearby sites, are used only if required.



Time series of daily means temperature, dew point, wind speed, total global and direct radiation, and annual percentiles (red lines) to assess consistency, for NH#60 Tullamarine.



## Appendix B Finkelstein-Schafer Statistic

The statistic for closeness of a month's data to the mean distribution is:

$$FS_{ym}^x = \frac{1}{n} \sum_{d=1}^n |D_{ym}^x(X_d) - D_m^x(X_d)|$$

where

$X_d$  is the value of parameter  $x$  on day  $d$

$D_{ym}^x$  is the distribution of parameter  $x$  in month  $m$  of year  $y$  (black, Figure 18)

$D_m^x$  is the combined distribution of parameter  $x$  in month  $m$  (red, Figure 18)

$n$  is the number of days in month  $m$  of year  $y$  with valid data.

An advantage of the F-S statistic is that, as a mean in probability space, it is dimension-free. Thus, it is directly comparable between different physical measures, so that a weighted sum of the F-S statistics for several quantities correctly reflects their specified importance without the need for prior normalisation. The weightings in this work are listed in Table 4.

The weights  $w_x$  are used in the obvious way to compute the combined F-S statistic of each year  $y$  for month  $m$ :

$$FS_{ym} = \sum_x w_x FS_{ym}^x$$

Note that the F-S statistic can be computed even for months with missing data for some days, and such months still contribute sensibly to the combined distribution functions and to the sorted set of weighted F-S values. Months with some missing data are thus still of value in establishing what is 'typical', but at the stage of selecting years for each month of the TMY we omit any with whole days missing for any parameter.

### Closeness to Long-term Mean or Median

The next step in the prescription of Marion and Urban (1995) is to select the 5 months with lowest combined F-S score, and rank them in order of "closeness of the month to the long-term mean and median". They do not say how they compare these two measures, nor how they weight them for the different parameters as both mean and median are expressed in physical units so would require some normalisation.

We explored several techniques for applying Step 2 of Marion and Urban, such as scaling the means by standard deviation and the medians by interquartile range, weighting both measures equally and then by the weights for each parameter. Our preferred technique, for consistency with Step 1, is to simultaneously compute a 'signed' F-S value defined, with the same notation as previously, by:

$$FSs_{ym}^x = \left| \frac{1}{n} \sum_{d=1}^n (D_{ym}^x(X_d) - D_m^x(X_d)) \right|$$

Referring to Figure 18, the true  $FS$  measures the mean absolute deviation of a month's distribution function (DF) from the combined DF, but a curve lying entirely above or below the reference curve can

score equally with one that crosses it. In contrast, *FSs* is smallest for a curve that lies equally above and below the reference and will consequently have a median close to the overall median.

The *FSs* values have the further advantages that they can be computed simultaneously with *FS* and weighted in the same way, they are again independent of physical units, and skewness of the underlying distribution is accommodated. Using *FSs* or other measures made only small changes to the order of preference for selected years.

Although in the end we did not use them in the selection process, means and standard deviations of daily values within a month were computed for all parameters. They provide a useful visual check of the results. Each TMY2 comes from 120 plots like Figure 18 (12 months x 10 parameters); all merged into F-S statistics which are not easy to review. Instead we show, in Appendix C, several examples of monthly means and standard deviations of solar radiation, temperature, humidity, and wind speed, with the selected months highlighted.

For convenience of comparison, the same scales are used for the corresponding plots in Appendix C, though this does put some data points off scale. Months chosen for inclusion in the TMY should be central for both mean and standard deviation, and this for all four variables. This objective is not fully achievable; the most typical months for mean radiation might be extreme for its variability, or for temperature or wind, for example. Appendix C shows that the TMY2 procedure produces reasonable results.

### **Persistence of High or Low Values**

In their Step 3, Marion and Urban (1995) prescribe that “persistence of mean dry bulb temperature and daily global horizontal radiation are evaluated by determining the frequency and run length above and below fixed long-term percentiles.” They use both terciles (33<sup>rd</sup> and 67<sup>th</sup> percentiles) for temperature, and the lower tercile for radiation. Applying the persistence criteria to candidate months from Step 2, they exclude “the month with the longest run, the month with the most runs, and [any] month with zero runs.” The implication of this description is that the most and least persistent of just the candidate months are excluded, without reference to whether those months are more or less persistent than usual for the long-term record. If, for example, all five months are more persistent in weather patterns than the long-term average, then surely the least persistent of those five should be preferred.

Marion and Urban (1995) are also less than clear what constitutes a ‘run’, but two consecutive values in the same tercile (high, medium, or low temperature; or low radiation or not) seems to be the criterion. This gives three separate run measures, and the question of whether they are to be tested separately or in combination. Do few runs for high temperature compensate for many runs of low radiation? With some difficulty interpreting the prescription, we adopted the following technique.

Histograms of sequential days within the above terciles are computed, and their cumulative sum gives the distribution function of run lengths of each type, analogous to Figure 18. The combined distribution

of run lengths enables evaluation of each month's distribution, as previously, with an FS-type statistic, *FSr* say.

$$FSr_{ym} = \frac{1}{10} \sum_{l=1}^{10} \sqrt{l} |N_{ym}(l) - \bar{N}_m(l)|$$

$$N_{ym}(l) = \sum_t w_t N_{ym}^t(l)$$

where

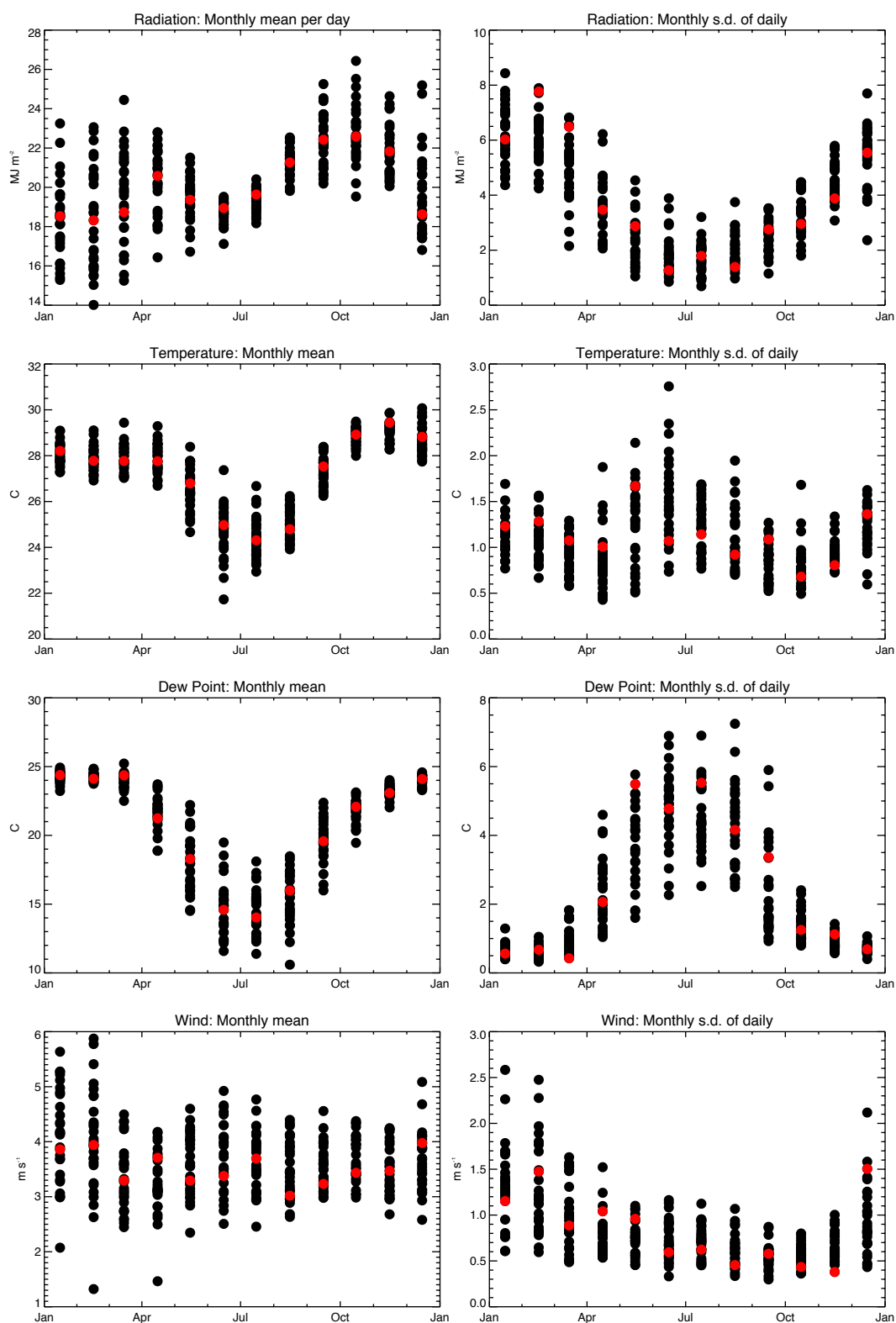
$N_{ym}^t(l)$  is the cumulative number of runs of length  $l$  in month  $m$  of year  $y$  for test  $t$  (parameter and tercile criterion)

$N_{ym}(l)$  is the weighted sum of the  $N_{ym}^t(l)$

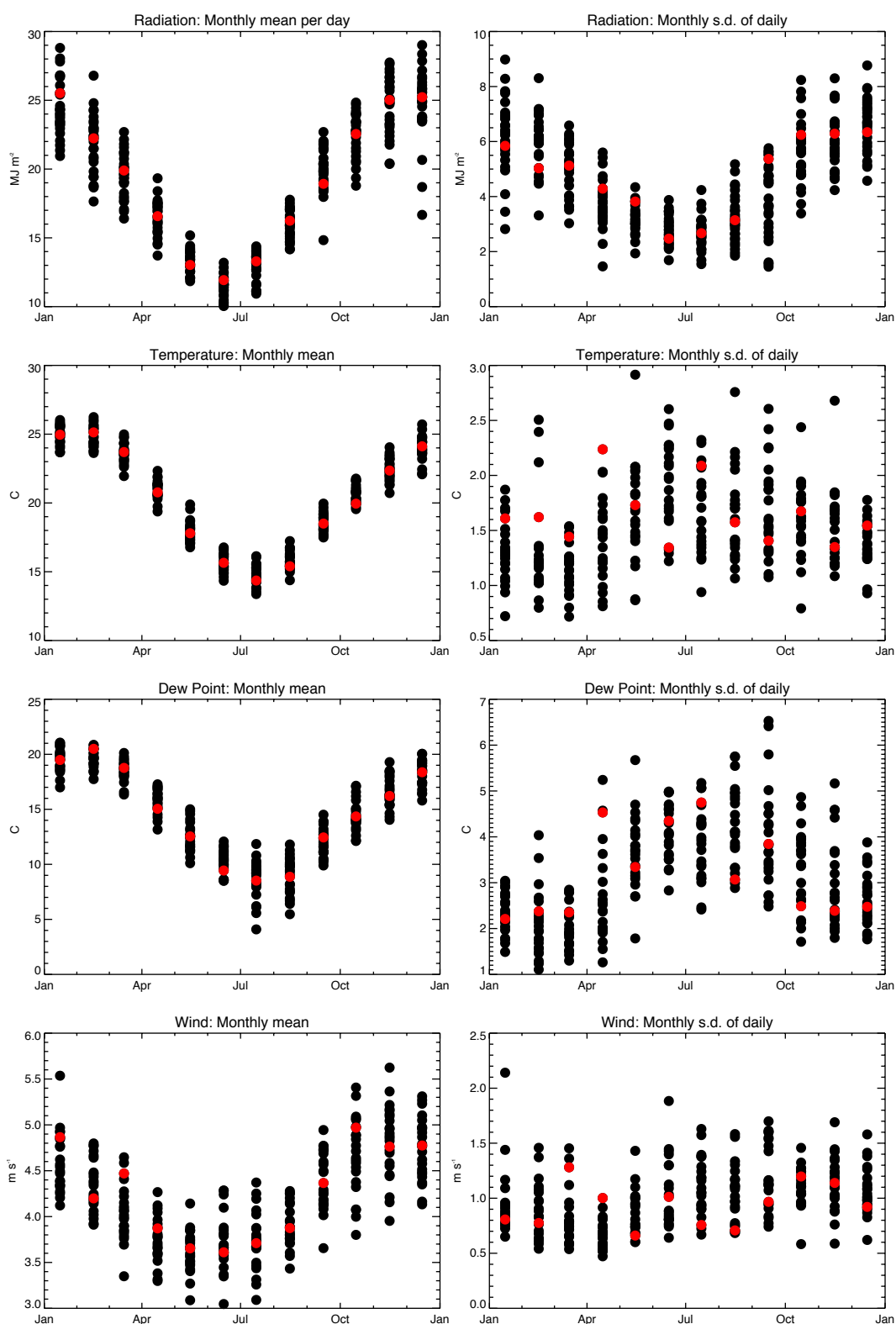
$\bar{N}_m(l)$  is the mean of  $N_{ym}(l)$  across all years.

For similarity to the earlier weightings for the 10 parameters, we separately considered runs of low global or direct radiation, and then with equal weightings  $w_t$ . The distribution of these *FSr* statistics across all years at several sites shows a long tail of high values in less than about 10% of cases. Selection of TMY-month years was thus restricted to below the 90<sup>th</sup> percentile for *FSr*.

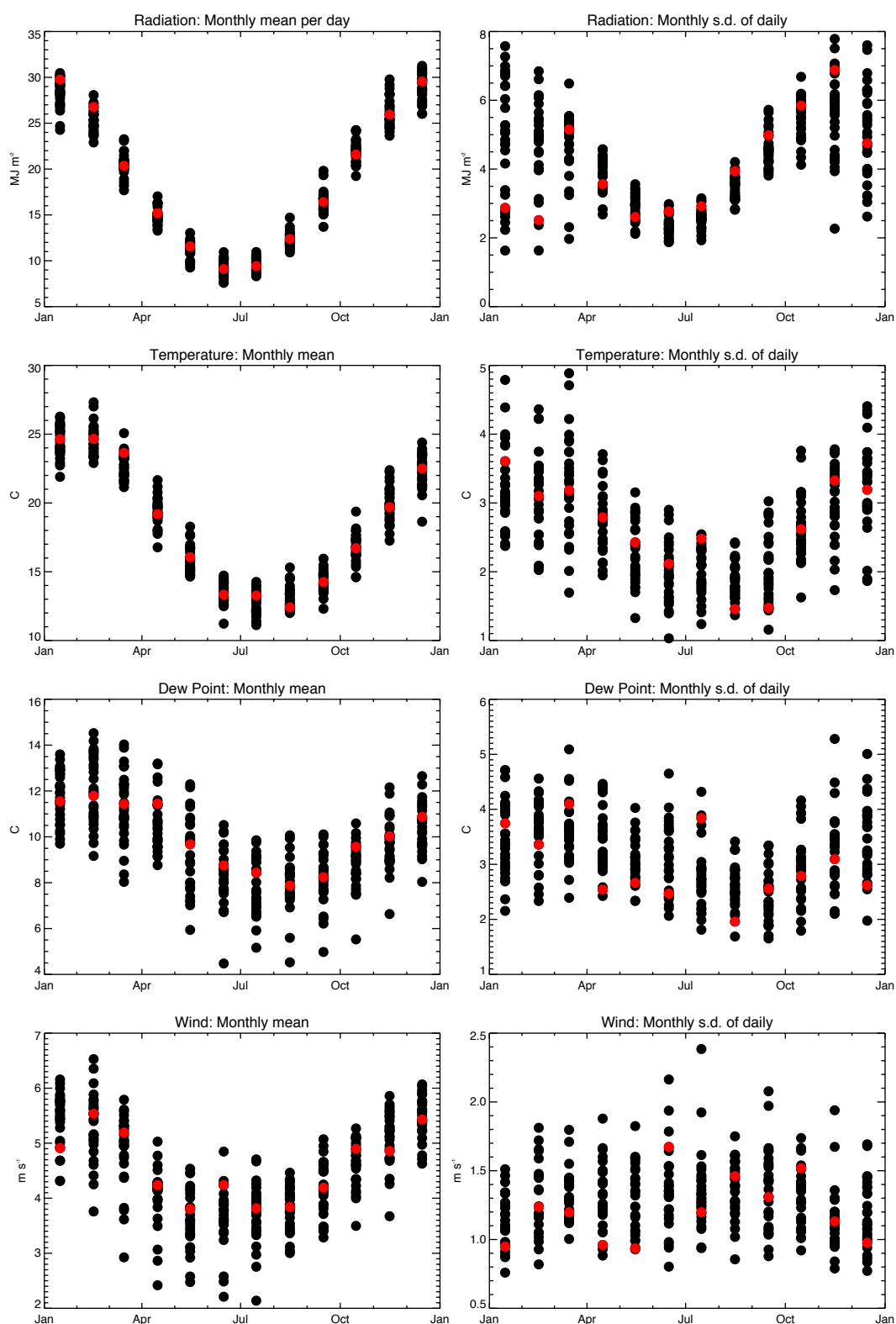
## Appendix C Monthly Mean and S.D. of $G$ , $T$ , $T_d$ , $W_s$



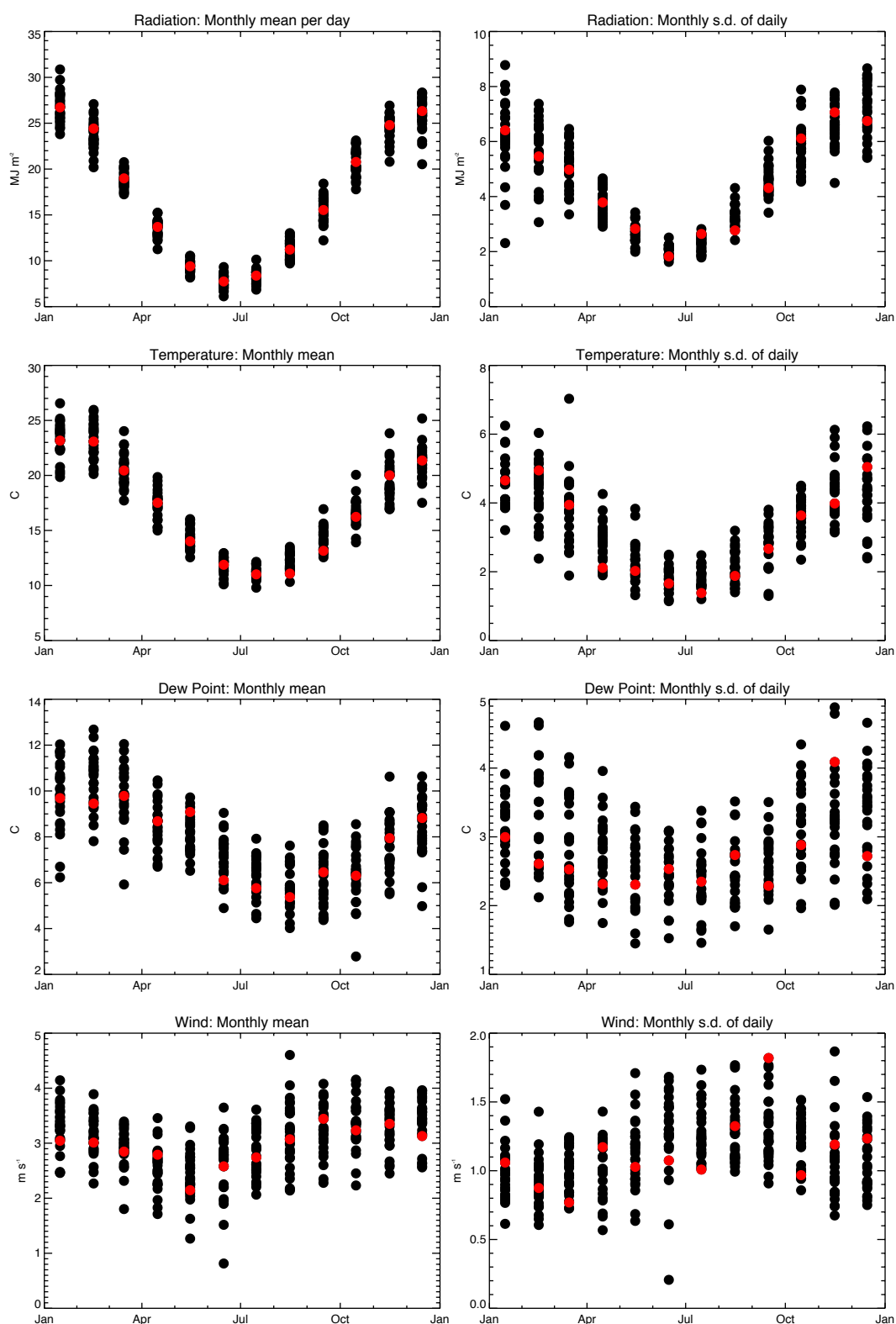
Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#01 Darwin.



Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#10 Brisbane.

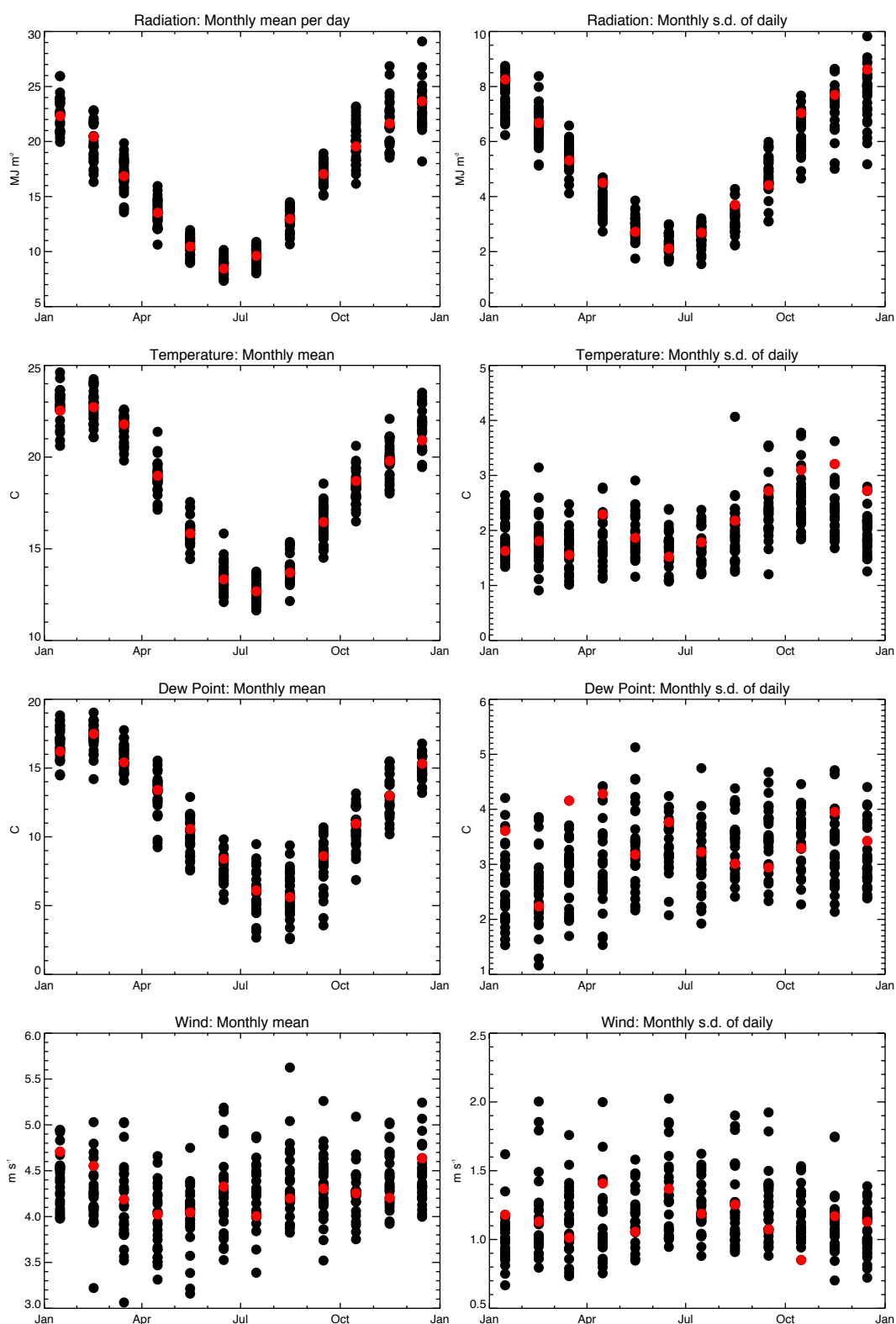


Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#13 Perth.

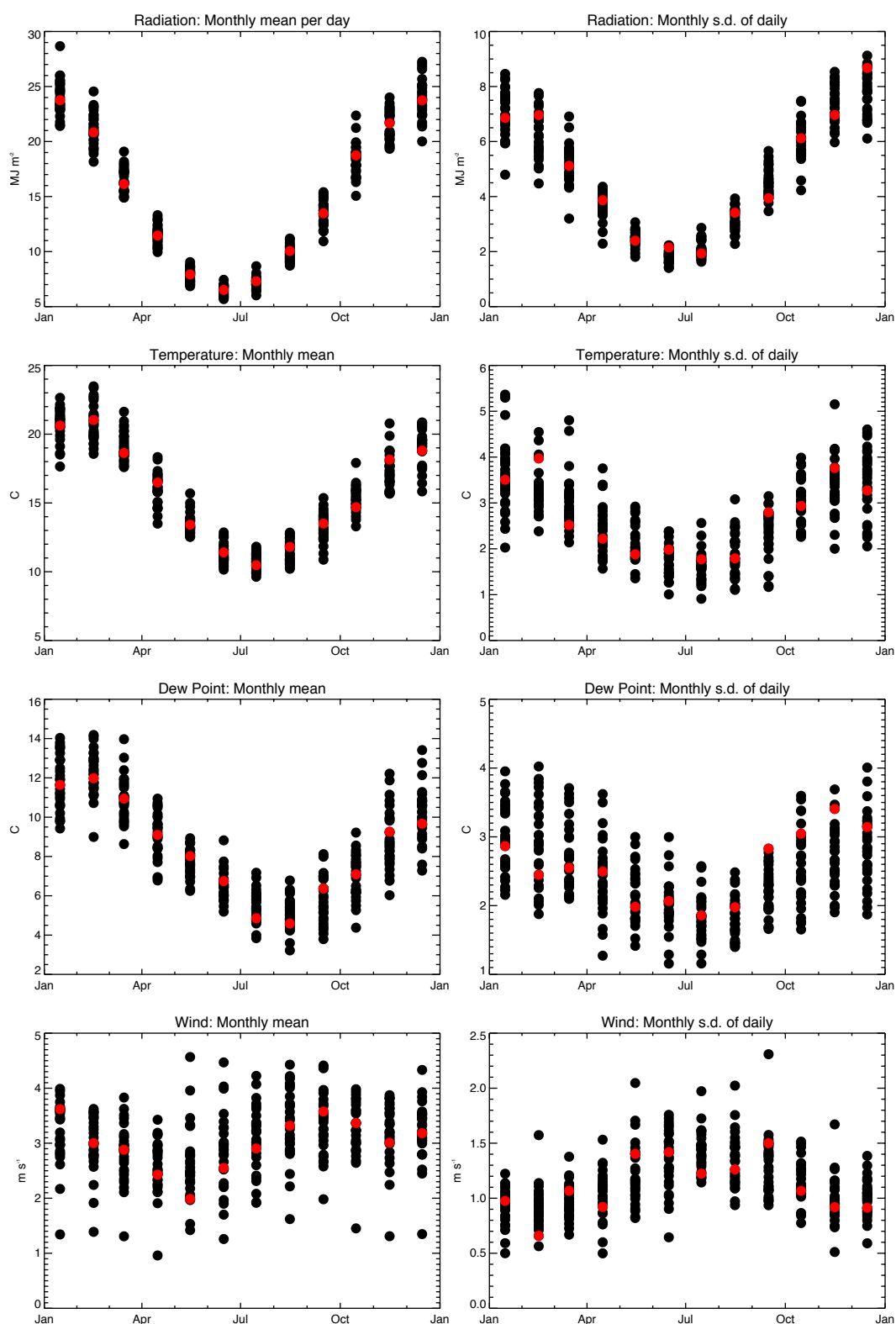


Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#16 Adelaide.

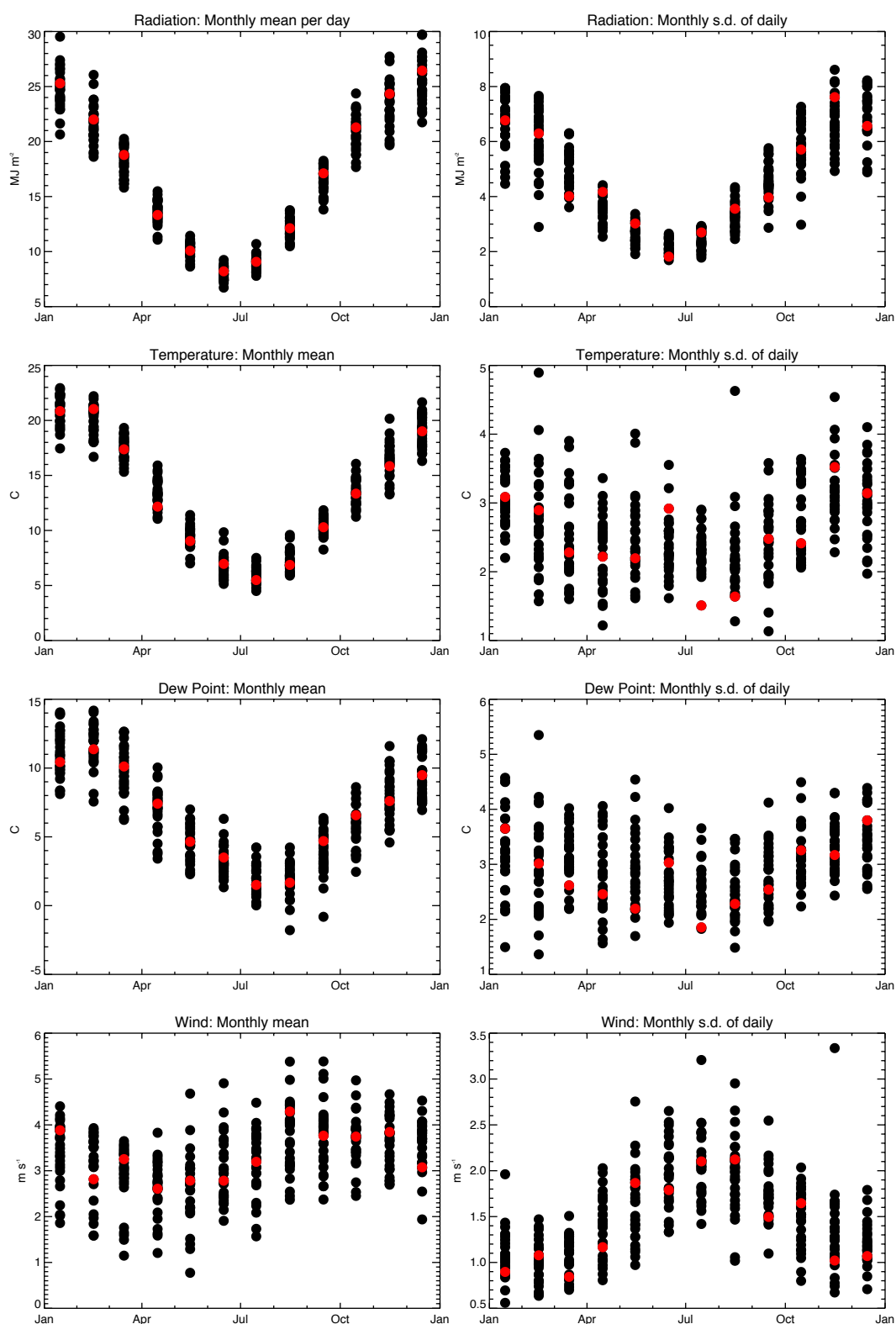




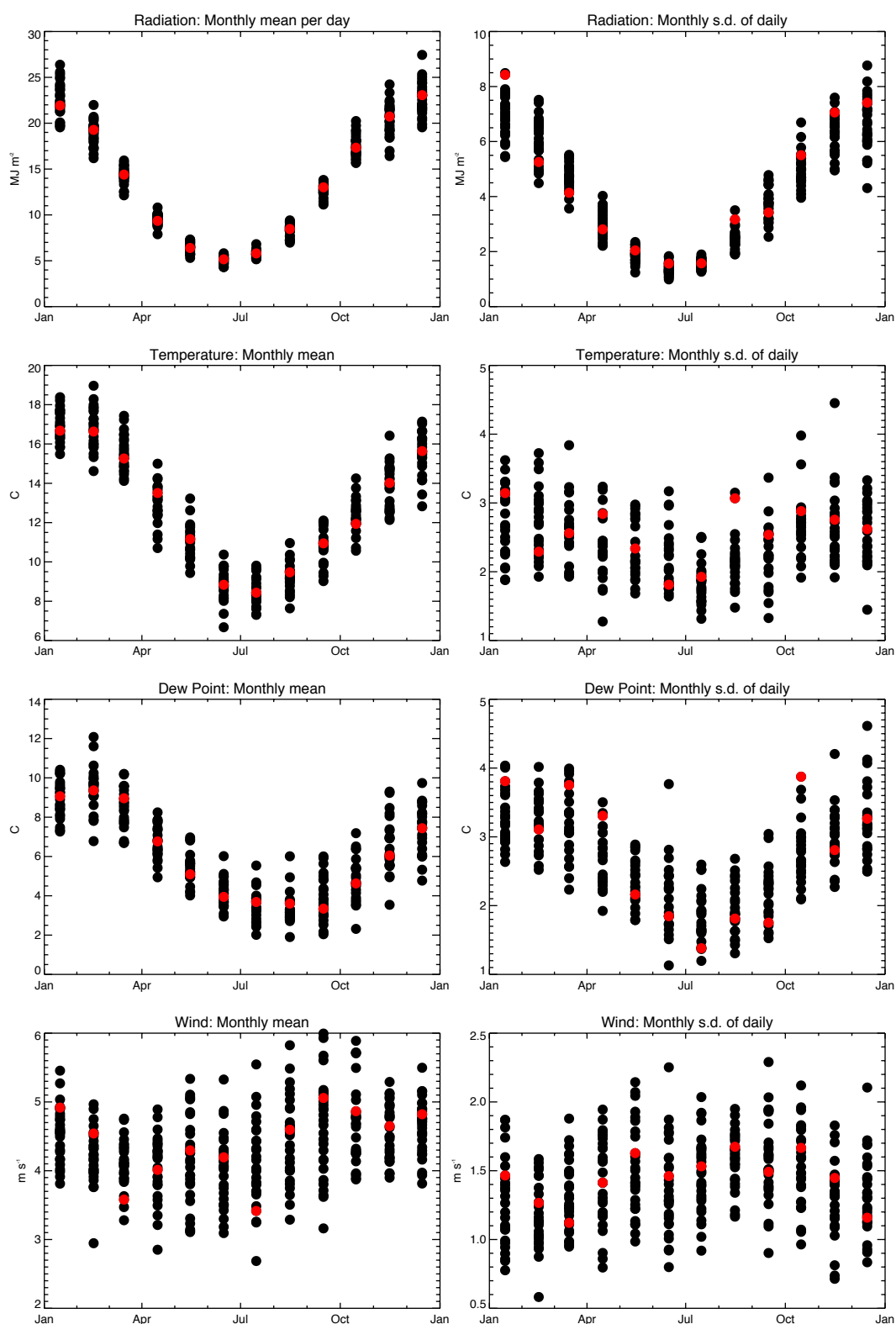
Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#17 Sydney.



Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#21 Melbourne.



Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#24 Canberra.



Monthly mean and s.d. of global irradiance, temperature, dew point, and wind speed for all months, with RMY selection shown in red, for NH#26 Hobart.