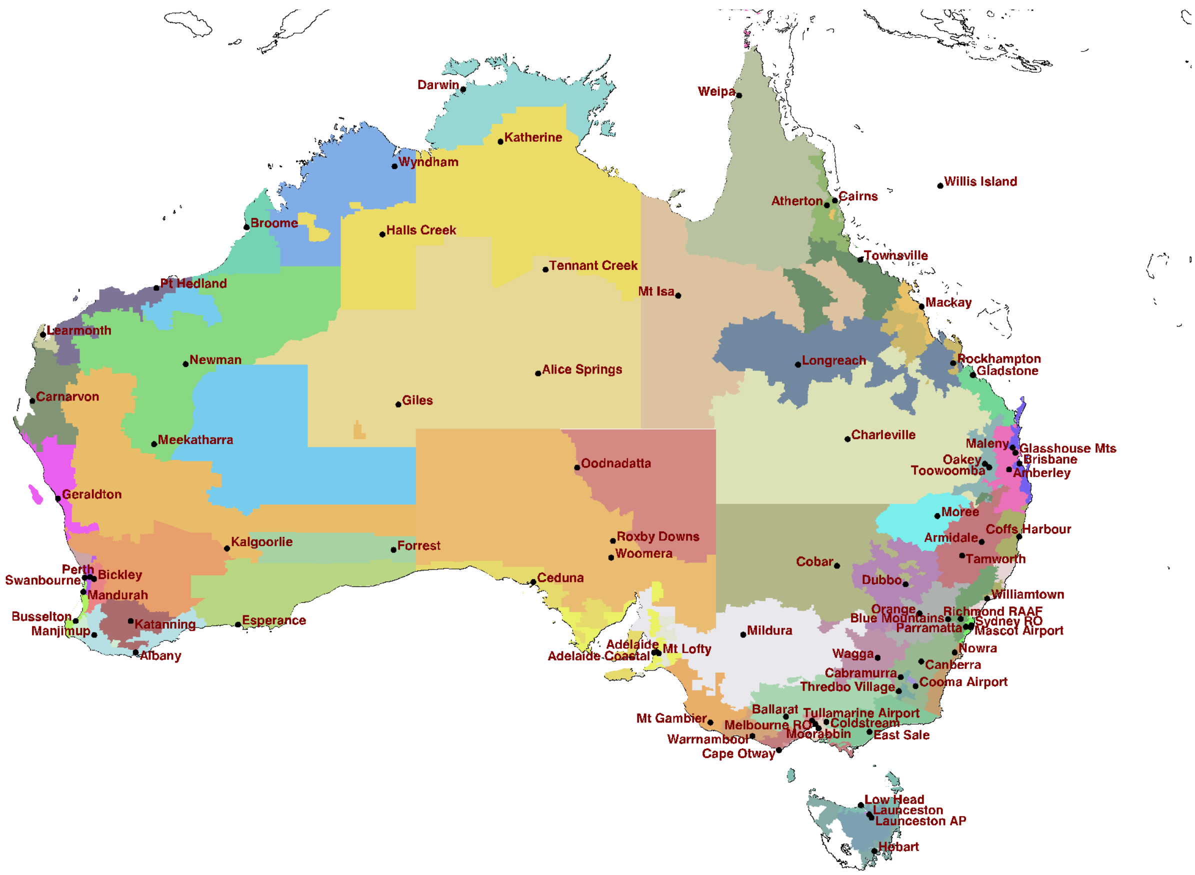
Creation of NatHERS 2016 Reference Meteorological Years

Including Maleny and Christmas Island



Prepared for Australian Federal Government
The Department of Environment and Energy

April 2017

Prepared by:
J Ben Liley

For any information regarding this report please contact:

J Ben Liley
Atmospheric Scientist
+64-3-440 0427
ben.liley@niwa.co.nz

State Highway 85, Lauder Central Otago

NIWA CLIENT REPORT No: 2017103WN
Report date: April 2017
NIWA Project: DCE16601

Quality Assurance Statement		
	Reviewed by:	David Pollard
	Formatting checked by:	P Allen
	Approved for release by:	Dr Andrew Laing

Cover picture: See Figure 2.

Contents

Executive Summary

The Reference Meteorological Years (RMYs) have been revised and updated for the Nationwide House Energy Rating Scheme (NatHERS). Unlike past revisions, which were based on Australian Climate Data Bank (ACDB) data files, this revision derives directly from half-hourly and synoptic data. To reflect more recent climate, and to make best use of improved instrumentation and data recording, we use data only since 1990, but now extending to the end of 2015.

In addition to the 82 sites in the 2011 version of NatHERS, an additional tropical island site has been created to use the high-quality radiation data from Cocos (Keeling) Island instead of Willis Island, for which no radiation data are available. The new site XI, #83, is labelled for the nearby more populous Christmas Island.

Estimates of global (*G*) and direct (*R*) solar radiation for the period 1990-2015, obtained by the Bureau of Meteorology from satellite data, are available for the entire Australian continental land mass, and interpolated to inshore islands. For 16 representative sites, high-quality ground-based measurements of *G*, *R* and the diffuse irradiance (*F*) are available for varying periods since 1990, including for Cocos Island since 2004. These data have been used as applicable, aggregated to hourly values from the original 1-minute resolution.

Algorithms for quality assurance and control, as detailed in this report, have been applied. Spline fits to mean sea level pressure across the continent have been derived to fill gaps or replace erroneous values of station pressure. Temperature and dew point have similarly been interpolated spatially where gaps are too long for just temporal interpolation. Outlying values, either in magnitude or in hour-to-hour change, have been marked as missing and the resulting gaps filled in the RMY files.

Extended periods of bad data were excluded from both the RMY selection and the statistical analysis that establishes typical patterns, so the RMY selection process is restricted to just the best data, where they cover a decade or more.

Previously, RMY files for NatHERS were produced in three sets, labelled RMYA, -B, and -C. These used different weightings of the Finkelstein-Schafer (F-S) statistics in the selection for typicality. In this revision, we retain only the RMYA weightings, as used internationally. The full time series are available, and may be of value for other uses of the data, especially for the derivation of 'typical' extremes, rather than just representative conditions. Anomalous values are not corrected in these time series, but data quality codes highlight improbable values.

We have developed an objective process for the delineation of climate zones. It uses data from synoptic stations to measure their similarity to adjacent NatHERS zones, with the same F-S statistical technique by which the RMY months are chosen. The technique is illustrated with respect to the Maleny zone, producing credible results but implying that higher-resolution data are needed.

Applied to islands off the Queensland coast, but therefore without radiation data, the new technique seems less conclusive, but suggests that continental files are valid to around 100 km offshore. It also demonstrates that the new Christmas Island zone is as representative as Willis Island for offshore tropical islands.

1 Background

Australia's Nationwide House Energy Rating Scheme (NatHERS) includes climate data in the form of Reference Meteorological Years (RMYs) to represent the country's different climate zones. The RMY data are deployed in the same file format as the Australian Climate Data Bank (ACDB), as developed originally by CSIRO (e.g., Chen 2016).

Previous revisions of the NatHERS RMYs have started with an update of the ACDB, expanded both in time and by the addition of new sites. An update by Energy Partners (2008), for the then Department of Climate Change and Energy Efficiency (DCCEE), extended coverage to the end of 2007 for a total of 41 years at most sites, and added 11 new zones to the 69 represented in NatHERS. The resulting ACDB dataset was later corrected and updated (Liley 2013) to the end of 2011. One of the extra zones in Queensland, labelled Maleny (MN, #73), was poorly represented by its nominal site. That dataset was relabelled as Glasshouse Mountains (GM, #82), and a new dataset for Maleny was created. The update also added a dataset for Busselton (BU, #81) in WA.

The present revision departs from previous work by starting afresh with Bureau of Meteorology data. Mostly, the data are from Automatic Weather Station (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 have been added where the record for previous sites was inconsistent or discontinued, but otherwise the first 82 sites are as in the previous revision.

Because the satellite-derived solar radiation data, as described below, are not available for the Willis Island station used to represent offshore locations, a new station has been created for Christmas Island (XI, #83). It is also not covered by the satellite-derived data, but as the nearby Cocos Island is a station in the international Baseline Surface Radiation Network (BSRN), it has solar radiation data of the highest quality.

Climate data in the RMYs used by NatHERS are hourly records of:

- Air temperature
- Moisture content / Wet bulb temperature
- Pressure
- Wind speed
- Wind direction
- Cloud cover
- Global (horizontal) solar irradiance
- Diffuse irradiance
- Direct radiation (on a sun-tracking surface)

Derivation of RMYs is based on the method for TMY2s described by Marion and Urban (Marion & Urban 1995) with the weightings suggested therein. The RMY designation, rather than TMY (Typical Meteorological Year), denotes that the 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.

The NIWA review (Liley 2010) of ACDB files had found many instances where meteorological values were anomalous, such as spikes in temperature by 10 °C for a single hour, or values that are extremely improbable for the site. Objective statistical tools to detect such anomalies were suggested in that report, and implemented in the subsequent work by NIWA (Liley 2013). They have been further refined and applied in the present work.

The most significant finding in the 2010 review was that techniques previously used to estimate solar radiation values for those sites and times without good measurements produced unphysical values in many instances. In particular, the estimates did not exhibit the strong relationships between global irradiance (G) as a fraction of its clear-sky value (G_c) and the diffuse (F) and direct (R) components, also as fractions of G_c, that are found in measurements from reasonably clean (low aerosol haze) sites. These relationships are very apparent in measurements from those (previously 16, now 20) NatHERS sites with radiation measurements of high quality, as demonstrated in the NIWA report.

A past issue with the radiation data, apparent for the sites with full *GFR* datasets, was the distinction between Mean Solar Time (MST), used by the Bureau of Meteorology before 1993, and True Solar Time (TST) as used for half-hourly irradiance values since then. In the 2012 update of NatHERS data, TST-aligned data were interpolated to MST as used for other meteorological variables in NatHERS. For the current revision, all site radiation data were obtained at 1-minute resolution, allowing hourly values to be calculated as expected by the Chenath engine in NatHERS software (Chen 2016).

For all other sites, and any periods without ground-based radiation data, we rely on estimates of G and R derived by the Bureau of Meteorology from geostationary satellite measurements. The time series of G and R for the Australian continent extend from 1990 to the present, and for the present work we used data to the end of 2015. The data as supplied from the Bureau of Meteorology include some values that are clearly erroneous from visual inspection. Error detection and correction for the satellite files developed for the previous NatHERS revision has been implemented again here.

A further question addressed in the new work is how adequately the set of NatHERS climate zones, whether 69, 82, or 83 of them, represent the country. This question is addressed with specific reference to the region around Maleny in Queensland. We also assess how far offshore around the Queensland coast the zone data can be used before the tropical island site of Willis Island (WS, #31) or the new Christmas Island (XI, #83) provides a better model.

2 Climate data time series

2.1 Locations

There are 69 sites currently used by NatHERS, based on those in the revision of ACDB to 2004. They are spread across the Australian continent to represent the range of climate zones, with higher density in populous areas to serve the predominant needs of building and energy system simulation. Locations of the sites are shown in Figure 1, with the then suggested boundaries for the zones represented by the enclosed sites.

All of these 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 completed records after temporal or spatial interpolation.

Most sites record atmospheric pressure, though it is relatively unimportant directly except as an indicator of meteorological change. Air pressure is a required variable to convert between the different representations of atmospheric water vapour; wet bulb temperature, dew point, relative humidity, and absolute moisture as represented in ACDB. Wind speed and direction are critical variables and for sites where they are not measured data from a sufficiently comparable nearby station are substituted.

Solar radiation is measured at only a few sites. In the past, many sites and time periods had only daily totals of G, whether measured or estimated from satellite observations, but NatHERS requires hourly values of G, F, and 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 choice of 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.

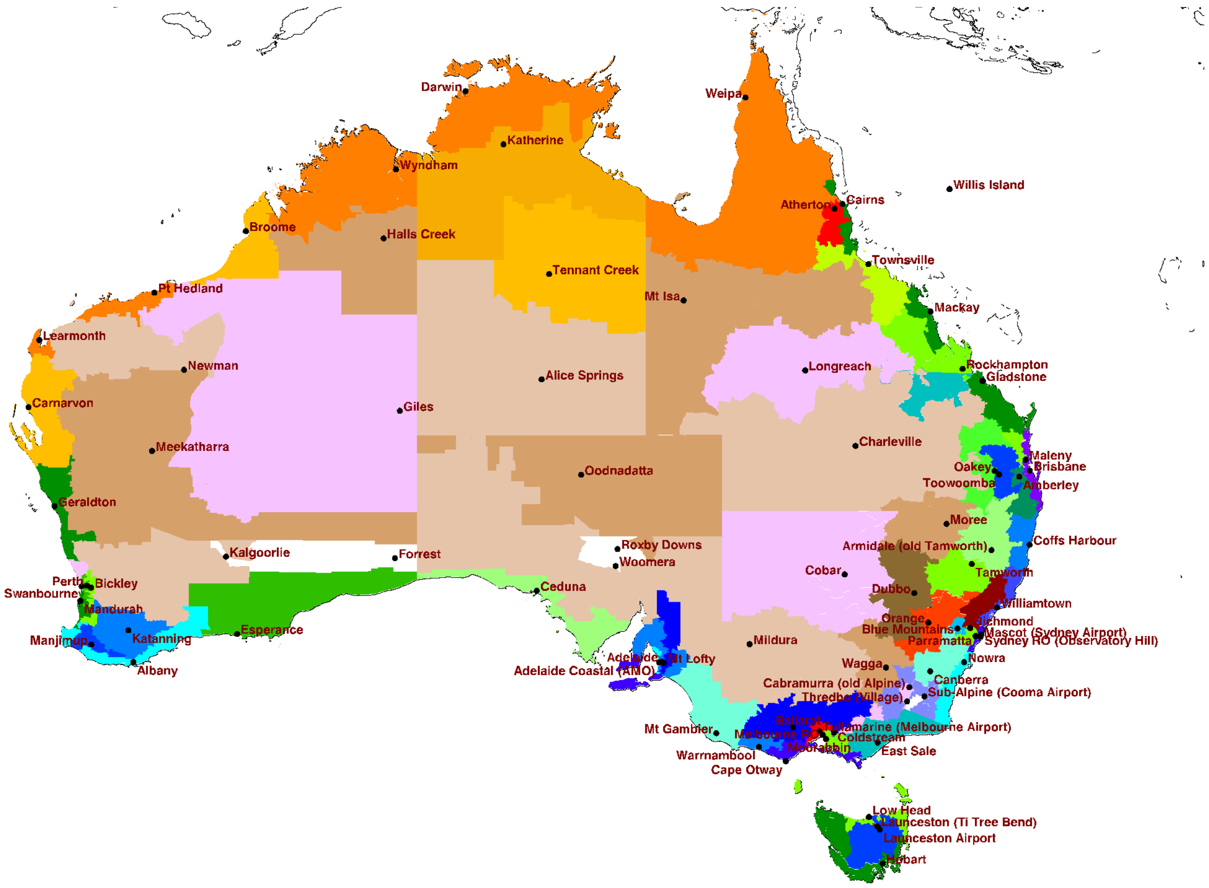


Figure 1. **Locations of 80 ACDB 2008 reference sites, and associated Climate Zones.**

As of 2016, NatHERS still designates just 69 Climate Zones, now redefined as shown in Figure 2 together with 82 sites for which climate data files have been developed.

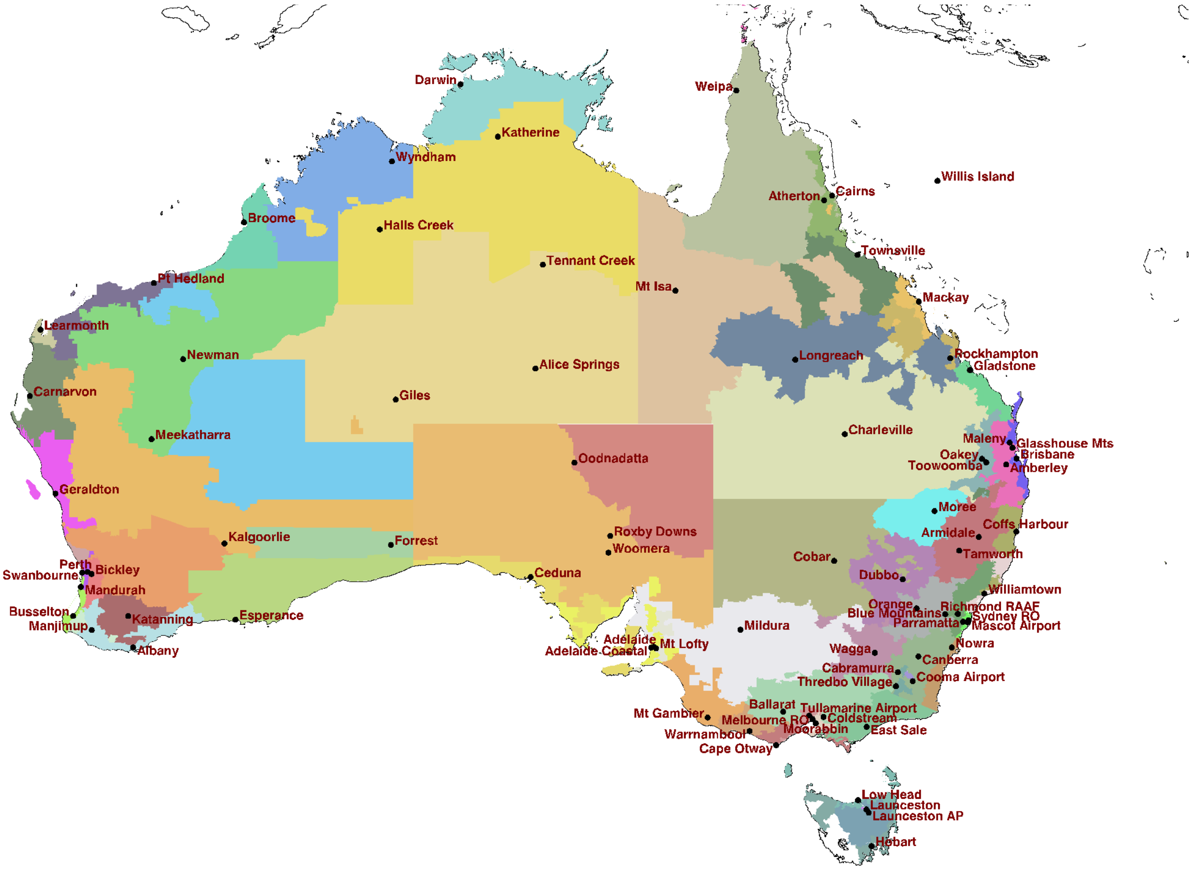


Figure 2. Locations of 82 reference sites, and the 69 NatHERS 2011 Climate Zones.

Table 1. **Locations of representative sites for 83 NatHERS climate zones: Zone follows ACDB zone coding; NH is the NatHERS zone number; Post is the current post code; Prev. is the previous post code for that site, if different; Time is relative to UT; CZ is the BCA climate zone; 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	Prev.	State	Alt.	Long.	Lat.	Time	CZ	BoM	WMO	Since
CZ0101	DA	Darwin	1	820	800	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	850	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	4883	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	4810	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	4000	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
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	860	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	870	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

Zone	L2	Name	NH	Post	Prev.	State	Alt.	Long.	Lat.	Time	CZ	BoM	WMO	Since
CZ0401	GI	Giles	41	872	6438	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	2650	NSW	213.0	147.457	-35.158	10.0	4	72150	94910	1990
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	6530	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	6000	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	Williamtown	15	2318	2300	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	2150	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	5000	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

Zone	L2	Name	NH	Post	Prev.	State	Alt.	Long.	Lat.	Time	CZ	BoM	WMO	Since
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	2541	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	5240	SA	685.0	138.709	-34.978	9.5	6	23842	95678	1990
CZ0607	TU	Tullamarine (Melbourne Airport)	60	3045	3020	VIC	118.8	144.832	-37.666	10.0	6	86282	94866	1990
CZ0608	CS	Coldstream	80	3770		VIC	83.9	145.409	-37.724	10.0	6	86383	94864	1995
CZ0609	ME	Melbourne RO	21	3053	3000	VIC	32.2	144.970	-37.807	10.0	6	86071	94868	1990
CZ0610	MG	Mt Gambier	61	5291	5290	SA	69.0	140.774	-37.747	9.5	6	26021	94821	1990
CZ0611	MR	Moorabbin	62	3194	3189	VIC	12.7	145.096	-37.980	10.0	6	86077	94870	1990
CZ0612	SE	East Sale	22	3851	3852	VIC	8.2	147.132	-38.116	10.0	6	85072	94907	1990
CZ0613	WR	Warrnambool	63	3275	3280	VIC	71.4	142.452	-38.287	10.0	6	90186	94837	1990
CZ0614	OT	Cape Otway	64	3238	3220	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	2600	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	3350	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	7250	TAS	5.0	147.122	-41.419	10.0	7	91237	94969	1990
CZ0708	LU	Launceston Airport	68	7258	7120	TAS	168.4	147.214	-41.549	10.0	7	91311	95966	1990
CZ0709	HO	Hobart	26	7004	7000	TAS	51.4	147.328	-42.890	10.0	7	94029	94970	1990
CZ0801	CM	Cabramurra (old Alpine)	25	2720	2629	NSW	1482.4	148.378	-35.937	10.0	8	72161	95916	1990
CZ0802	TH	Thredbo Village)	69	2627	2625	NSW	1380.0	148.304	-36.492	10.0	8	71041	95908	1990

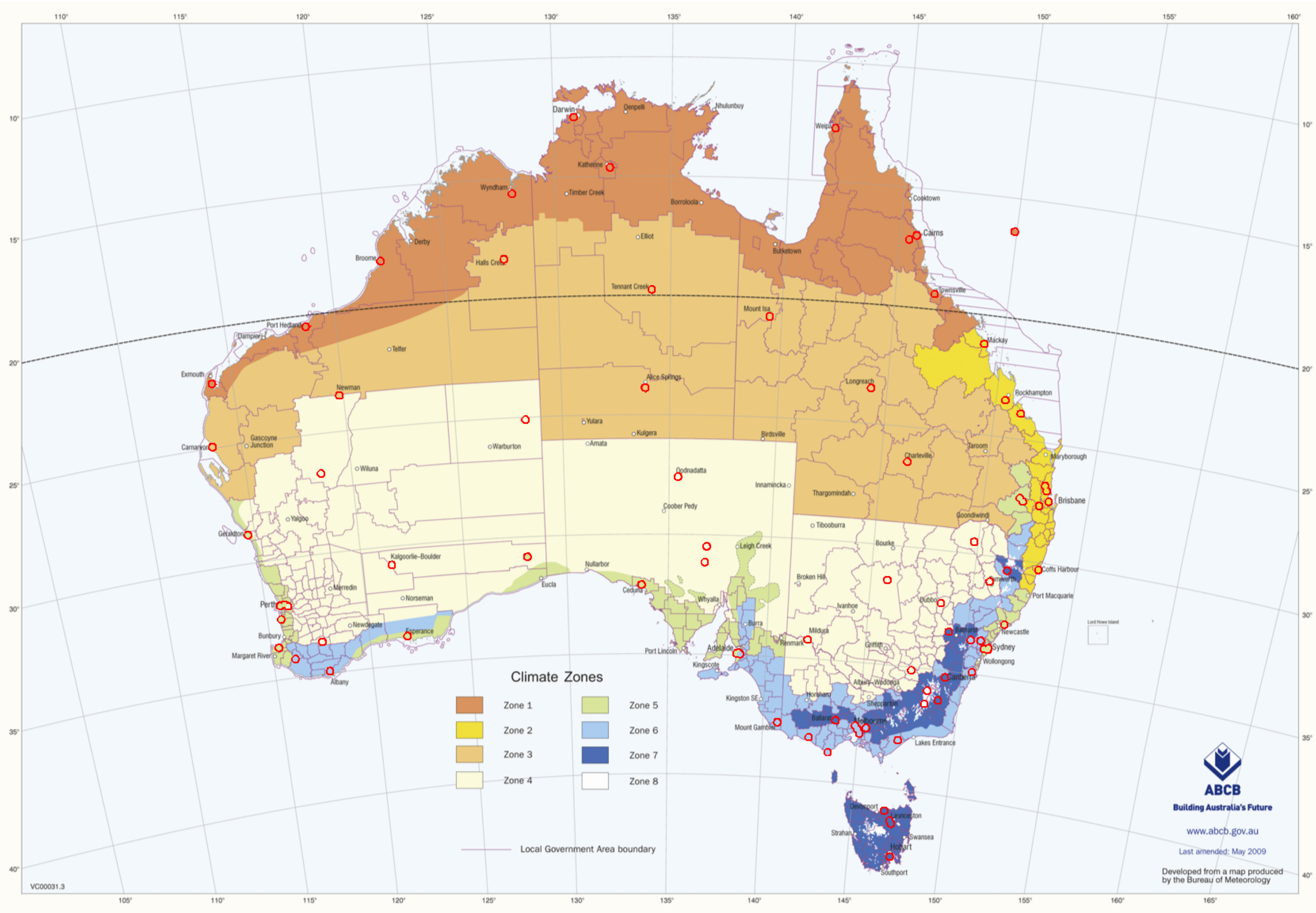


Figure 3. **Locations of 82 reference sites within the ABCB Climate Zones. The interior of the red circle shows the nominal Climate Zone for that site. The new Christmas Island (XI, #83) site, represented by Cocos (Keeling) Island at 12.2° S, 96.8° E, lies outside the map at upper left.**

2.2 Adjustments

Determining the boundaries of the NatHERS Climate Zones was not part of the work that NIWA undertook in the project reported here. That has been managed by NatHERS in consultation with State representatives and local expertise. Nonetheless, the extent to which climate data can be used to inform that process and help resolve uncertainty in where boundaries should lie is considered later in this report.

Use of the radiation data in NatHERS requires calculation of sun angles, which depend on location and time. Assignment of the correct satellite-derived radiation data relies on precise locations of all representative sites within its 0.05° x 0.05° grid. Consequently, all locations have been checked and if necessary amended to the latest Bureau of Meteorology coordinates of the primary climate station for that climate zone. Table 1 lists the locations used for NatHERS 2016, and related parameters for each site, plus the two new sites introduced herein. Column NH denotes the ACDB zone number, extended from the 69 prior zones, and used here for NatHERS.

In the revision of site details, there were minor corrections to the altitudes of zones WS #31, AM #9, IS #39, CR #4, KA #44, BI #47, MI #27, OA #50, GE #12, WE #15, BU #81, AB #58, and OR #65. Longitude or latitude changed very slightly (by less than 0.1°) for zones GE #12, AB #58, OR #65 and TH #69; for the first two, this was because a different Bureau site was denoted as primary.

Column CZ in Table 1 denotes to which of the eight major Australian climate zones identified by the Australian Building Codes Board (ABCB) the NatHERS zone belongs. Columns Post and Prev. give the 2012 post code for each site and the post code in 2008. Many of them changed, and three sites (Alice Springs, Tennant Creek, and Giles) fall within the single post code 872, even though Giles is in a different ABCB climate zone.

With 83 sites and 69 NatHERS zones, it is inevitable that many zones now include more than one site. Nevertheless, as apparent in Figure 2, several NatHERS zones have no representative site within their boundary, and the reference site lies in another zone:

CZ0401 (GI, #41)	Giles	is in CZ0306 (AL, #6, Alice Springs)
CZ0402 (MT, #42)	Meekatharra	is in CZ0305 (NE, #40, Newman)
CZ0603 (MJ, #57)	Manjimup	is in CZ0605 (AB, #58, Albany)
CZ0801 (CM, #25)	Cabramurra	is in CZ0703 (CA, #24, Canberra)

The second numerical digit of the zone code is the ABCB Climate Zone, and it is unfortunate that three of these four sites are represented in NatHERS by a different ABCB zone. Figure 3 is copied from the ABCB web site, as listed on the figure, and shows the ABCB zones. The 82 NatHERS sites that lie within the map area are denoted by the red circles, with interiors coloured by the nominal ABCB zone as in Table 1, using the same colours. Apart from Forrest in WA, near the Great Australian Bight, the classifications of Table 1 agree with the boundaries in Figure 3.

2.3 New zone

The representation of small islands far from the Australian coast is problematical because radiation data from satellites are confined to land large land areas. The values can be extrapolated a short distance offshore, but more remote islands need a better representative. The Willis Island (WI, #31) site is currently used in NatHERS for this purpose, but it cannot be updated without radiation data, and even the historical radiation data are of uncertain provenance.

For this revision, data other than radiation were updated, and the RMY recalculated in case mean temperature, humidity or wind had changed. Unlike at all other sites, for which data only since 1990 have been used, the Willis Island time series back to 1977 has provided the basis of the RMY derivation.

To provide an alternative to Willis Island, this revision of NatHERS makes use of high-quality radiation measurements by the Bureau at Cocos (Keeling) Island. They are combined with other climate data from both Cocos and Christmas Island. Because of its larger population, the latter gives its name to the new site (XI, #83).

3 Climate data

3.1 Half-hourly data from AWS

The great majority of climate data for NatHERS have come from half-hourly records of Automatic Weather Stations (AWS) in the Bureau of Meteorology network. Of 665 AWS available sites, this NatHERS revision has made use of 99, because some sites lacked essential data for a substantial part of the period 1990 – present.

As noted above, this revision departs from previous practice by not updating the ACDB files, instead deriving the time series afresh from Bureau data. In part this is motivated by the interval of satellite-derived radiation data, but also because much of the early ACDB data were interpolated from 3-hourly or less frequent values, or else estimated by various means. The restriction to post-1990 values has greatly reduced the need for estimates, though some are still necessary.

3.2 Missing data

Atmospheric pressure data in ACDB were especially problematic, and the issue extended to new data obtained from the Bureau of Meteorology. Several of the AWS sites do not report air pressure, and it is also not reported in synoptic (3-hourly, twice daily, or less frequent) climate data, probably because the values would have little meteorological significance.

The datasets for NatHERS require atmospheric pressure for psychrometrics (humidity conversions). Measured station pressures, temperatures, and dew points for all sites were ingested and run through the validity checks described below. Pressures were converted to Mean Sea Level Pressure (*MSLP*), which varies smoothly and only slowly over continental scales. The *MSLP* values were spatially interpolated by kriging to other sites, and the resulting estimates converted back to station pressure to replace missing values. Interpolates for temperature and dew point were derived similarly, but because air temperature in particular varies over much shorter spatial (and temporal) scales the resulting values were devalued in the statistical procedure for RMY selection described later.

Although the spatial interpolation works well for hourly data, it can give anomalous results for synoptic values. The many more measurements at 0900, 1200, 1500, etc., than at adjacent times affects the relative weightings. Spatial interpolation alone thus resulted in unrealistic steps in pressure with time, and these affected other detection of errors. To avoid the problem, synoptic values were first interpolated in time.

3.3 Visibly anomalous data

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. For example, the above process of generating spatial interpolates was repeated no less than 18 times, with each one requiring visual rechecks of results.

Even in extensive review, 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 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 detection of 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. There can also be odd values that do not stand out except by comparison with adjacent values. We have previously found good discrimination of such events with the following technique, which was 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 Pt Hedland, we fitted a simple model of seasonally varying diurnal cycle, measured departures from the model, and smoothed them to a timescale typical for weather systems. Individual temperature differences from the result are represented in Figure 4 as a histogram with logarithmic ordinate. Points outside a fitted envelope are highlighted in red.

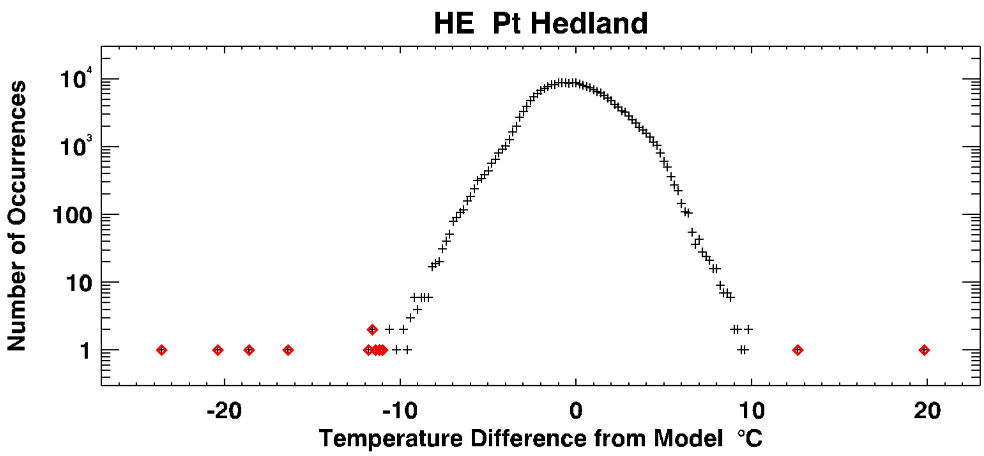


Figure 4. Histogram of hourly temperature difference from the smoothed mean, both relative to seasonal and diurnal cycles. The values highlighted in red are likely to be anomalous.

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. The five values between -12 and -10 °C, as selected by the algorithm, 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.

In Figure 5, the anomalous values are again highlighted in red. The plot shows that in this instance the respective months were not selected in RMYs, but the inclusion or exclusion of anomalous values still affects the statistical procedure for TMY month selection described later.

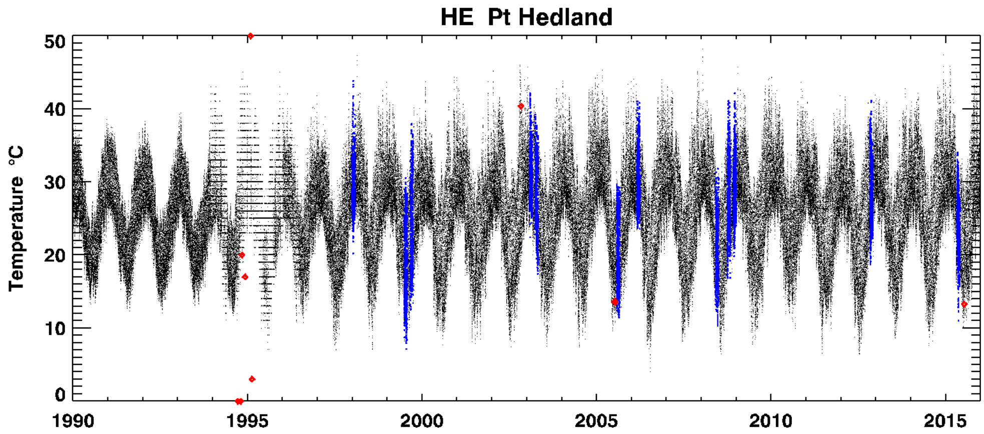


Figure 5. **Time series of hourly temperature, with RMY data highlighted in blue. Eleven values, highlighted in red, were found by the outlier algorithm to be anomalous.**

These anomalies are mostly just single hourly values, and after their removal there may still be sequences of two or more anomalous values. These are detected by an algorithm that uses the histogram of all hour-to-hour differences, as in Figure 6.

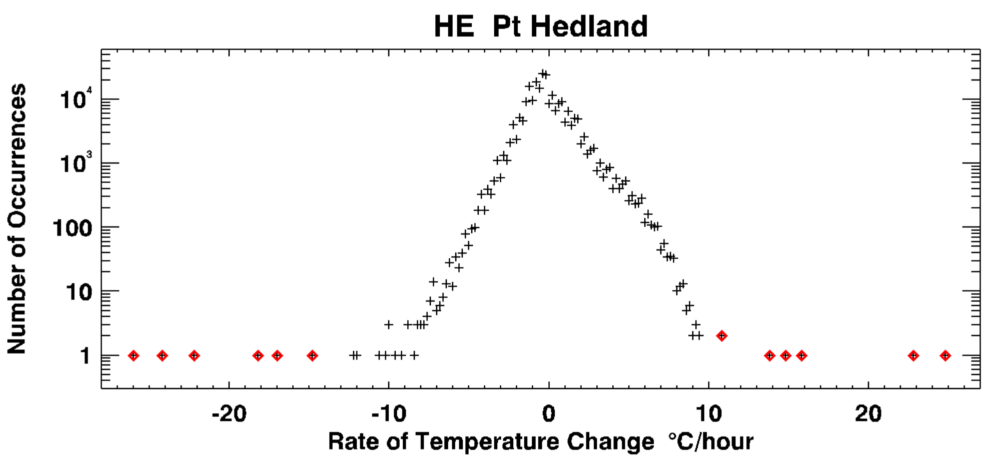


Figure 6. **Histogram of hour-to-hour temperature differences, highlighting in red those data points removed by the two algorithms.**

Any algorithm to detect anomalous values must meet two criteria; it needs to find (the great majority of) erroneous data, but it should accept genuine weather extremes. Wind data for northern, especially coastal, locations like Pt Hedland include incidents of gale force winds, probably associated with tropical cyclones. For this reason, it is difficult to establish limits in absolute value, but wind speed changes are still informative. The histogram of wind speed changes from hour to hour is shown in Figure 7, analogous to Figure 6 for temperatures. Red diamonds again denote atypical values of the differences according to the algorithm. Visual inspection of these 18 values confirmed that all are indeed dubious.

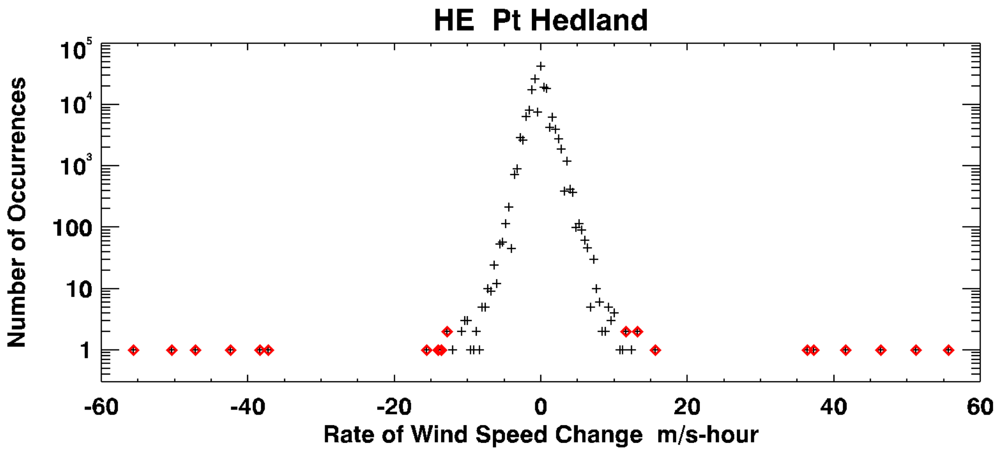


Figure 7. **Histogram of hourly wind speed change from previous value. The values highlighted in red, detected by algorithm, are considered anomalous.**

Figure 8 shows the corresponding time series, analogous to Figure 5. 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.

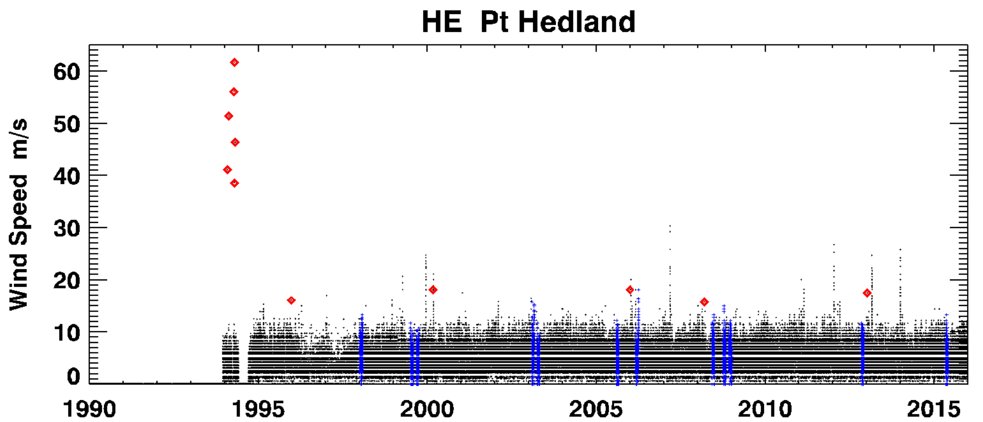


Figure 8. **Time series of hourly wind speed, with RMY data highlighted in blue. The values highlighted in red are anomalous, and the most extreme values are discounted, but regular incidents of high winds pass the test.**

The same analysis for pressure did not always avoid false positives, largely because pressure only changes slowly other than in exceptional circumstances. A classic example from our previous analysis of ACDB data for Darwin is shown in Figure 9, where an extreme pressure drop early on 25 December, 1974 was identified by algorithm as erroneous. As history records, it was very real.

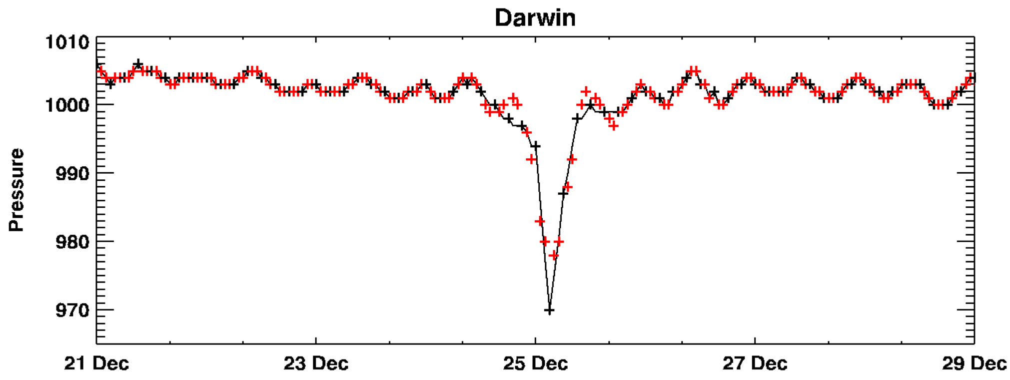


Figure 9. Time series of atmospheric pressure for Darwin, 21 – 29 December 1974 in ACDB data, with measurements in black and ACDB 2008 estimates in red.

Such incidents 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 added a further step of checking 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.

3.5 Drying of wet bulb wick

Extensive work, by a combination of automated detection and visual inspection, was required to diagnose and eliminate extreme values of absolute moisture at sites often hundreds of kilometres from the coast. The mechanism by which this occurs is illustrated in Figure 10 For Meekatharra, 430 km from the coast in Western Australia.

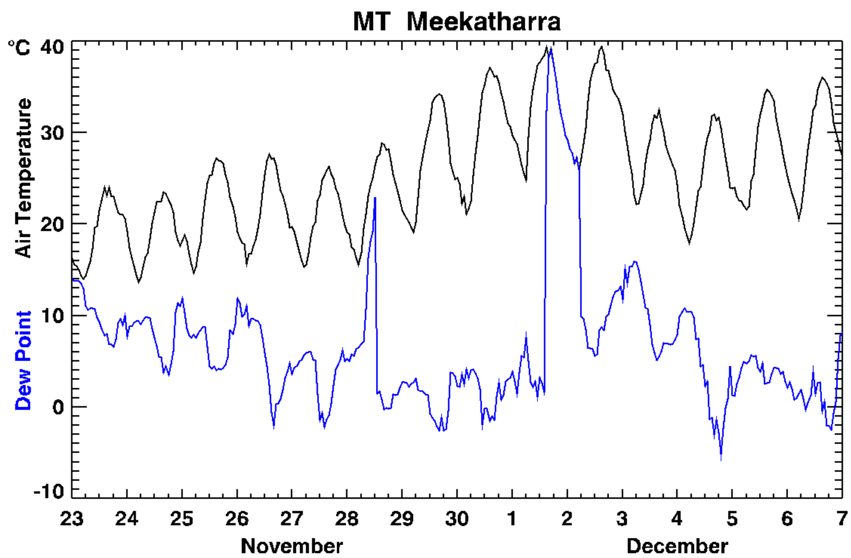


Figure 10. Hourly data from the Bureau of Meteorology, AWS site 7045, in 2008. Calculated absolute moisture is nearly 50 g kg-1 when air temperature approaches 40 °C at 100% relative humidity.

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 close to 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 dew point measurement 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 in recent Bureau data; perhaps more so if sites are visited less frequently. The algorithm developed to detect it uses a combination of criteria tested across a large number of data sets. We examined the distribution of dew points for all sites, and to the credible peak values we fitted a function

T_{dx} = 32.2 (°C) - 0.13 (°C/°S) x Latitude - 0.004 (°C/masl) x Altitude,

which ranges from 21.6 to 30.6 °C over the NatHERS sites. For dew points that exceed this value, or whose rate of increase or decrease exceeds 1.5 or -2 °C hr^{-1} respectively, the wet bulb wick is assumed to have dried. Those values are then marked as missing, to be filled by interpolation like other gaps in dew point data.

From visual examination of hundreds of incidents, our dry wick detection algorithm gave convincing results for the hourly data from the great majority of AWS sites.

4 Additional climate data

4.1 Sources

For continuity in NatHERS, the Bureau of Meteorology stations previously used for ACDB, as identified from the production report for ACDB 2008, were the primary source for the new time series. As noted earlier, two changes were Geraldton (GE, #12) and Albany (AB, #58), for which the representative Bureau sites were changed from 8051 to 8315 and from 9741 to 9999 respectively.

The great majority of primary sites were amongst Australia's 665 Automatic Weather Stations (AWS), reporting hourly values. Supplemental data from several manual recording sites amongst the 6880 open stations helped to extend coverage either in time or in measurement type, as there are often no cloud data from an AWS. Table 2 lists the stations used in addition to those identified in Table 1. In total 115 stations, of which 99 were AWS and 16 were synoptic sites, contributed data and were used in the spatial interpolation scheme for missing or anomalous Temperature, Dew Point, and Pressure.

Where only one climate variable is listed in Table 2, it was the reason for that site's inclusion. Where several variables are listed, the one marked with an asterisk was the reason for inclusion, but other values are also used to substitute missing or estimated values from the primary station.

Canberra Airport (70351) supersedes the discontinued station 70014.

Zones not represented by stations reporting hourly are Maleny (Bureau of Meteorology station 40121), Katanning (10916), Manjimup (94617), and Thredbo Village (71041). There are no hourly-reporting (typically AWS) stations near to the middle two. For Thredbo, the nearest AWS (71032) is at the top of the ski field (1950 m asl), but the NatHERS reference site 71041 is in the village at 1380 m altitude. The pressure data from 71032 are used directly, as the barometer is at 1367.9 m altitude.

In previous work for NatHERS 2011, drawing on ACDB 2008, Maleny (MN, #73) was based on climate data from Beerburrum Forest Station, Bureau of Meteorology station 40284, as the nearest AWS. In fact, that was a poor proxy, because of its proximity to the coast and low altitude of 48 m asl rather than 425 m asl as at the Tamarind St synoptic site (40121) in Maleny. Now, Maleny is represented by twice-daily temperature and humidity data from Tamarind St, interpolated to hourly with a surface fitted to surrounding AWS data. Wind data (unavailable for 40121) are from the Jimna Forestry AWS (40651) at 523 m asl, and radiation data are for the Tamarind St location. The Beerburrum AWS data are retained as the basis for a climate zone named Glasshouse Mountains (GM, #82).

For the other recent NatHERS zone, Busselton in WA, the primary Bureau station is the airport AWS (9603). There is little difference between it and the Busselton AWS (9569), except that the latter record has more missing data and station 9569 closed in November 2011.

4.2 Adjustment and spatial interpolation

Other data from the 71032 AWS on Thredbo ski field are adjusted to the correct altitude by lapse rate, and splined into the (typically twice daily) data from station 71041 in the village when possible. Wind data from the summit are reduced by a factor of 0.6 for compatibility with wind data from station 71041.

At Sydney Royal Observatory (66062), as addressed in the review of ACDB 2008, there were no wind data after April 1992, and the wind data from the beginning of 1992 are somewhat different from earlier data. There are good wind data from August 1990 to the present for Fort Denison (66022), but wind speeds are somewhat higher there than at the 66062 site. Rather than transform the Fort Denison data for compatibility, as we did previously, we have used them directly, so that the 66022 site serves as the new reference location for central Sydney wind.

Just as the ACDB data required interpolation of MSLP to replace invalid pressure data, the new 2008-2011 data from the Bureau of Meteorology contained gaps that needed to be filled, and there are no pressure data for Maleny, Bickley, Swanbourne, Blue Mountains, Mt Lofty, Launceston (Ti Tree Bend), Cabramurra (old Alpine), Thredbo (Village), and the new Glasshouse Mountains. Again, we created a spatially interpolated surface of MSLP, but also fitted surfaces of air temperature and dew point, adjusted by lapse rate to equivalent values at sea level. The interpolated surface for MSLP was used for the nine sites listed above, with adjustment back to station values by either measured station temperature or interpolated sea level temperature.

The interpolates of sea level temperature and dew point were adjusted back to station values by lapse rate, and either used directly for long gaps or splined into smaller gaps. Data gaps of less than three hours are effectively linearly interpolated by this procedure, while longer gaps obtain the area-average temporal variation for that time.

4.3 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.

Table 2. Additional Bureau of Meteorology stations used to extend ACDB to 2011. Data types listed alone, or marked with an asterisk, were the reason for using that station. Additional hourly data from AWS sites are used only if the primary station data (Table 1) are missing. If present, Alt P is the altitude of the barometer. 'Cloud' indicates cloud observations, 'Ceil.' denotes a measurement by ceilometer.

CZ	L2	BoM	Name	State	Alt.	Alt. P	Lat.	Long.						AWS
1	WY	1013	Wyndham	WA	11.0	16.0	-15.487	128.125					Cloud	
	AT	31108	Walkamin Research Station	QLD	594.0		-17.135	145.428					Cloud	
2	GL	39326	Gladstone Airport	QLD	16.6	16.9	-23.870	151.221	Wind	T	T_{dew}	P	Ceil.*	Y
	MN	40651	Jimna Forestry	QLD	523.0		-26.664	152.461	Wind	T	T_{dew}			Y
4	RX	16065	Andamooka	SA	76.0		-30.449	137.169					Cloud	
	BI	9021	Perth Airport	WA	15.4	20.0	-31.927	115.976	Wind	T	T_{dew}	P	Ceil.*	Y
	KT	10579	Katanning Comparison	WA	310.0	311.0	-33.689	117.555					Cloud	
5	GE	8051	Geraldton Airport Comparison	WA	33.0	35.0	-28.795	114.698	Wind	T	T_{dew}	P	Ceil.	Y
	SW	9172	Jandakot Aero	WA	30.0	30.7	-32.101	115.879	Wind	T	T_{dew}	P	Ceil.*	Y
	MD	9887	Mandurah	WA	21.0	22.0	-32.521	115.750	Wind	T	T_{dew}	P		Y
		9194	Medina Research Centre	WA	14.0		-32.221	115.808					Cloud	
	PA	66137	Parramatta North (Masons Dr)	NSW	55.0		-33.792	151.018	Wind	T	T_{dew}		Cloud*	
	SY	66022	Fort Denison	NSW	2.0		-33.855	151.225	Wind					Y
	AD	23034	Adelaide Airport	SA	2.0	8.2	-34.952	138.520	Wind	T	T_{dew}	P	Ceil.*	Y
	BU	9569	Busselton	WA	3.9		-33.655	115.319		T	T_{dew}			Y
6	BL	63039	Katoomba (Murri St)	NSW	1015.0		-33.712	150.309					Cloud	
	MJ	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	
	AB	9741	Albany Airport Comparison	WA	68.0	69.0	-34.941	117.802	Wind	T	T_{dew}	P	Ceil.	Y
	ML	23878	Mount Crawford	SA	525.0	525.5	-34.725	138.928	Wind	T	T_{dew}	P		Y
		23733	Mount Barker	SA	363.0		-35.064	138.851					Cloud	
	CS	86071	Melbourne Regional Office	VIC	31.1	32.2	-37.807	144.970	Wind	T	T_{dew}	P	Cloud*	Y
	ME	86068	Viewbank	VIC	66.1	66.4	-37.741	145.097	Wind	T	T_{dew}	P		Y
		86338	Melbourne (Olympic Park)	VIC	7.5	7.5	-37.826	144.982	Wind	T	T_{dew}	P		Y
	WR	90171	Cashmore Airport	VIC	80.9	81.5	-38.315	141.471	Wind	T	T_{dew}	P	Ceil.*	Y

CZ	L2	BoM	Name	State	Alt.	Alt. P	Lat.	Long.						AWS
7	AA	56037	Armidale (Tree Group Nursery)	NSW	987.0		-30.524	151.672					Cloud	
	OR	63231	Orange Airport Comparison	NSW	948.0		-33.381	149.123	Wind	T	T_{dew}		Cloud*	
		63303	Orange Airport	NSW	944.7	945.3	-33.377	149.126	Wind	T	T_{dew}	P	Ceil.	Y
	CA	70351	Canberra Airport	NSW	577.1	577.6	-35.309	149.200	Wind	T	T_{dew}		Ceil.	Y
	BA	89105	Lookout Hill	VIC	965.0		-37.282	143.247	Wind	T	T_{dew}			Y
	LD	91126	Devonport Airport	TAS	8.0	9.5	-41.170	146.429	Wind	T*	T_{dew}	P	Ceil.	Y
	LT	91311	Launceston Airport	TAS	166.9	168.4	-41.549	147.214	Wind	T	T_{dew}	P	Ceil.*	Y
	LU	91104	Launceston Airport Comp.	TAS	166.0	178.0	-41.450	147.203	Wind	T	T_{dew}	P	Ceil.	Y
	HO	94008	Hobart Airport	TAS	4.0	27.4	-42.834	147.503	Wind	T	T_{dew}	P	Ceil.*	Y
8	CM	72043	Tumbarumba Post Office	NSW	645.0		-35.778	148.012					Cloud	
	TH	71032	Thredbo AWS	NSW	1957.0	1367.9	-36.492	148.286	Wind*	T	T_{dew}	P		Y

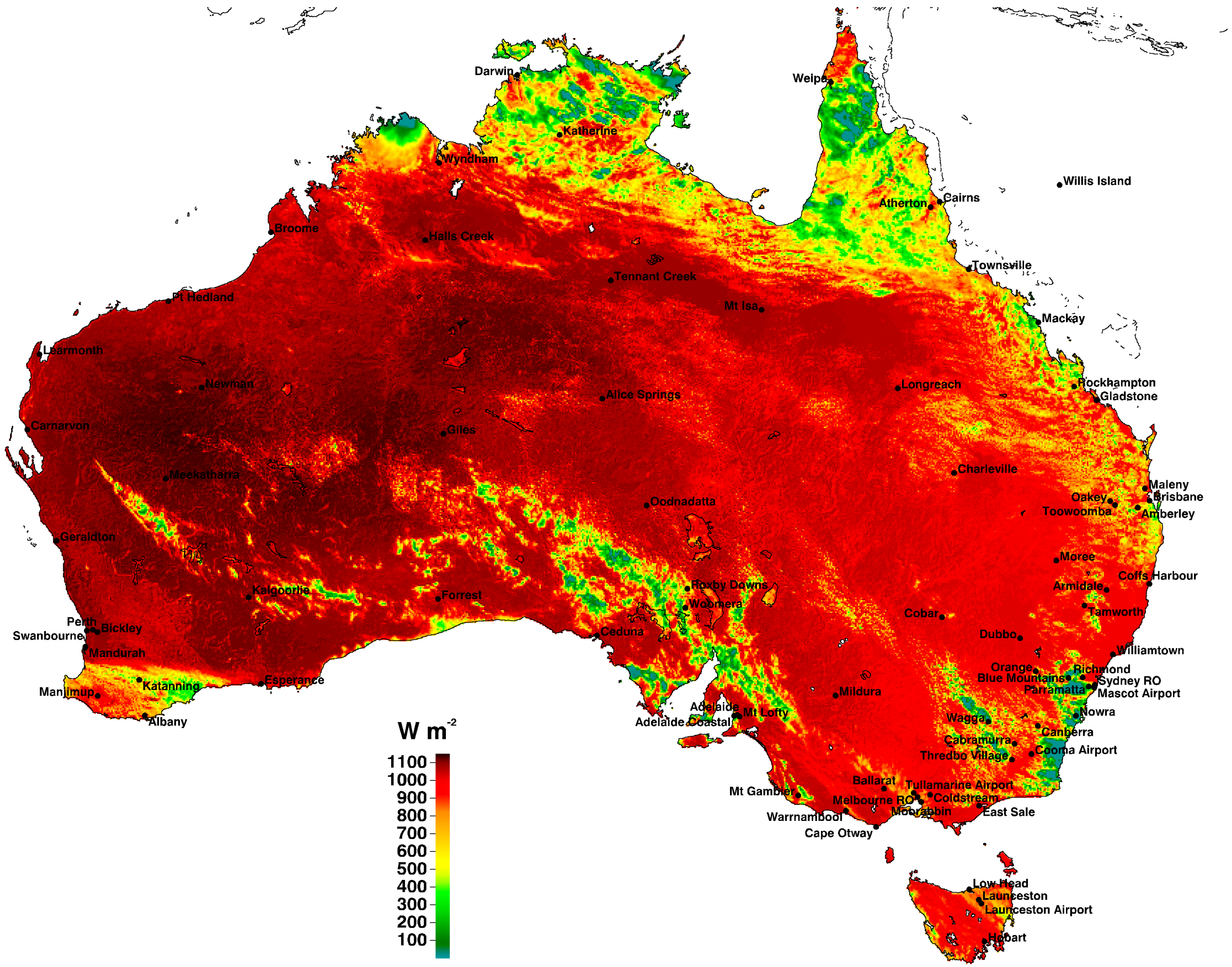


Figure 11. Global irradiance at (nominal) 0300 UT, 1 January 1998, derived from satellite data by the Bureau of Meteorology, overlaid with NatHERS reference sites.

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° x 0.05° (~ 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.

4.4 Wind data

A significant problem discerned in this update concerned the wind data for many sites. Plots like those shown 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 updates. In some cases, the anemometer was moved from near ground to an 8-m mast as assumed for NatHERS data. In other instances, there were gaps in data records that had to be filled from a nearby climate station.

As with the Sydney Royal Observatory, the issue is 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 this revision, new data sources were needed for several sites, including Melbourne.

Whether from an obviously different period at the same site, or a new site with different wind climate, wind speeds were transformed for reasonable consistency of the median (50th), 90th, and 98th percentile, as illustrate in the plots of Appendix A.

5 Solar radiation

5.1 Irradiance data

In the literature, the ratio of global irradiance, *G*, to its value at the top of the atmosphere has been called the *clearness index*. A more useful quantity for the following analysis is the *clear-sky index* (*CSI*), which expresses *G* as a fraction of G_c, the global irradiance for the same solar position in a clear sky. The development of Typical Meteorological Years for New Zealand (Liley et al. 2008) used values of *G* from 18 sites (amongst over 100) recorded in the NIWA Climate Database, and an algorithm to infer *F*, and downward direct (*R*cos(*Z*), where *Z* is solar zenith angle) irradiance from *CSI*. The algorithm is illustrated with New Zealand data in Figure 12, where the abscissa is *CSI*, and the ordinates show its partition into diffuse and down-welling direct components.

5.2 Measured *G*, *F* and *R*

There are measurements of *G*, *F* and *R* at 20 Bureau of Meteorology climate stations, 16 of which are representative sites in NatHERS. The measurements, to Baseline Surface Radiation Network (BSRN) standards, are among the best such data in the world. The *GFR* plot for Tullamarine (TU) is shown in Figure 13, including all values for the first 41 years for which *G*, *R* and *F* were measured.

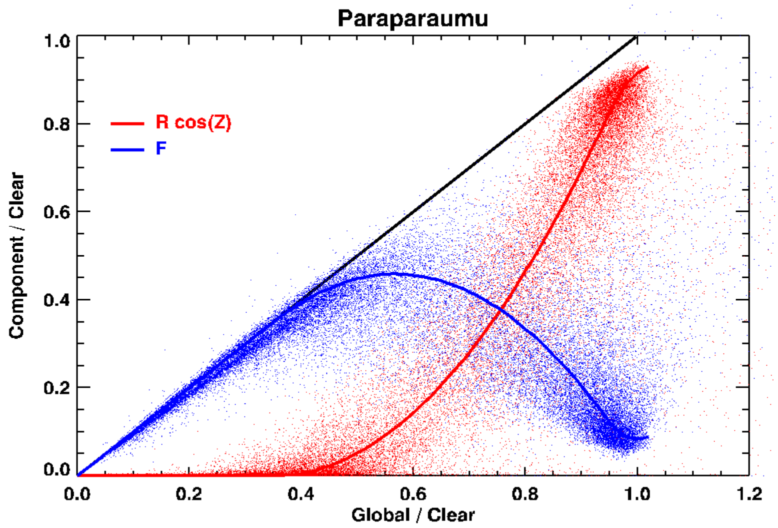


Figure 12. **Measured diffuse and down-welling direct components of global irradiance, all expressed relative to model clear sky irradiance. The scattergram shows all hourly data for Paraparaumu over 21 years that satisfy the consistency criterion that the sum of components is equal to global within about ± 5% of seasonal midday values. The fitted curves are used to predict *F* and *R* from measured *G*.**

Reassuringly, Figure 13 shows the same relationship as in Figure 12. There are still some erroneous data, where $F < 0.05\ G_c$ (blue points on or near *x* axis) or $F = G$ for high *CSI*. These are measurement anomalies (diffuse detector unshaded, or direct sensor misaligned) that can generally be filtered by requiring that the diffuse and direct components should (nearly) sum to measured global irradiance.

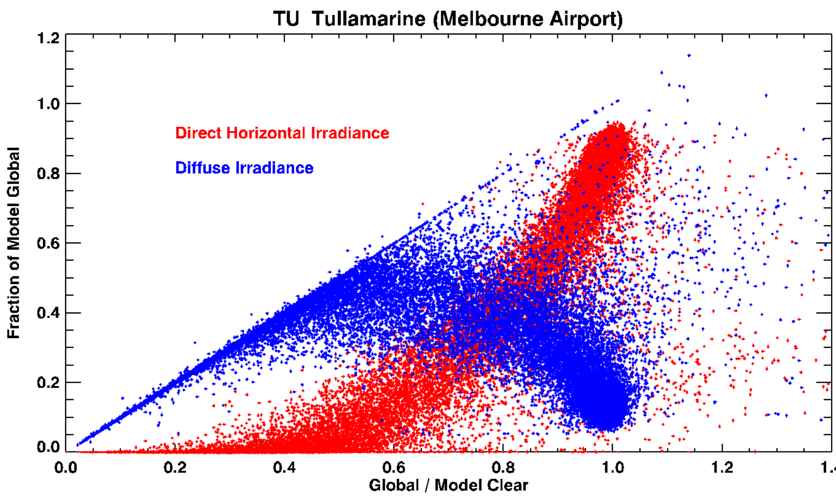


Figure 13. **Diffuse and down-welling direct fractions of global radiation for measured data at Melbourne Airport.**

In past use in NatHERS of ground-based radiation data, expressed as half-hourly values, there were anomalies even in the refined dataset regarding the times of observations. They are illustrated in Figure 14, which shows all the good global radiation data for Tullamarine in ACDB 2008 by day of year, highlighting the values for 09:00 and 16:00.

It is clear from Figure 14 that the data values for mid-September to early December, each year, are offset in time from the rest of the data. The same thing occurs at the other sites, but at different times of year and by different amounts. This anomaly arose because the ground-based radiation data after 1993 were supplied by the Bureau of Meteorology in half-hour increments of true solar time (TST); effectively sun-dial time, as the sun is due north (in the southern hemisphere) every day at noon in TST. The ACDB and NatHERS assume mean solar time (MST), for which the sun is on average due north at noon, but cycles about it over the year according to the Equation of Time.

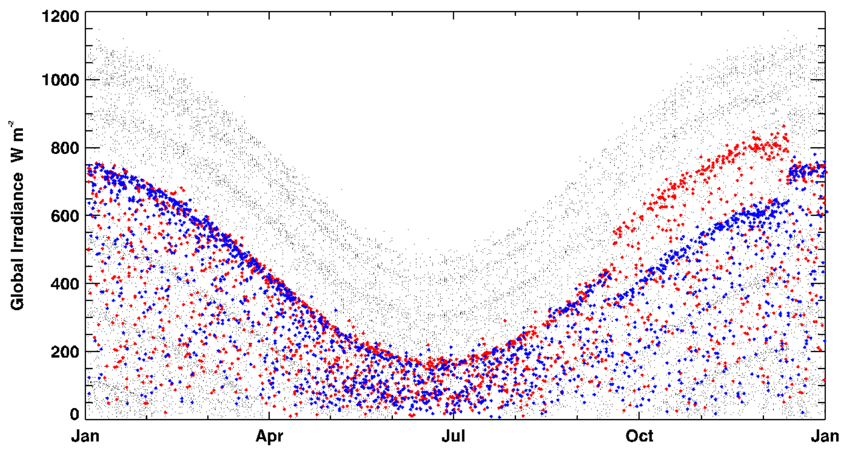


Figure 14. Good measured global irradiance data in ACDB 2008 for Tullamarine, in the period 1997-2007, by time of year. Grey points show all data, coloured symbols denote 09:00 (red) and 16:00 (blue) data, highlighting an apparent step in recording time.

For NatHERS 2012, the ground-based *GFR* measurements were shifted back to MST within each day, by interpolating the integral of daily radiation from one time scale to the other. This maintained the daily total, but reallocated cumulative radiant energy (joules per square metre) between hours to represent more accurately the average radiant power (watts per square metre) in each hour.

In the present revision, NatHERS 2016, all ground-based radiation data were obtained from the Bureau at 1-minute resolution, so that the hourly NatHERS values could be calculated explicitly in MST as expected by NatHERS software.

5.3 Radiation measurement quality

In our previous work for NatHERS, a different strategy was required for sites with only global irradiance measured, and not to current Bureau standards. Values of G less than 0.05 G_c (for solar elevation greater than 5°) were marked as erroneous; the first radiation data flag was set to 'missing'. Then, any values of R incompatible with the NIWA algorithm illustrated in Figure 12 and Figure 13 (outside the range of red points) were also marked as erroneous. In the derivation of RMYs, such missing values were replaced with estimates, but in the full time series the bad data remained, marked by data flags.

Because of the availability from 1990 of the radiation data described below, improvements in meteorological instruments over the decades, and both global and possibly local changes in climate, the present revision started afresh with data from the Bureau of Meteorology since 1990 and no reliance on past interpolation or estimates. The only exception was Willis Island (WS, #31), for which no satellite data are available. It is still based on hourly measurements of global irradiance apparently made there in the period July 1977 to May 1995.

5.4 Satellite-derived radiation data

As used in NatHERS 2012, a new product of research by the Bureau of Meteorology (Grant 2009) became available to fill the need for 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. As detailed in Section 5.5 of this report, we compared these data with measurements at the 16 sites where *F* and *R* were measured, and found good agreement. The satellite data are instantaneous, rather than hourly totals, but their distribution is comparable to hourly totals. To match this interpretation, and associated values of solar elevation and azimuth, the satellite-derived data have been interpolated to Mean Solar Time.

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. A complete set can be purchased on an external hard drive from the Bureau.

The data are provided at 0.05° x 0.05° resolution, the pixel size in Figure 11, which corresponds approximately to a 5-km grid. Analyses 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. 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 11.

Subsequent processing of satellite data by the Bureau of Meteorology extended the coverage back to 1 January 1990, and forward to near-present by quarters of the year, and largely filled the period from 1 July 2001 to 30 June 2003. Documentation from the Bureau 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.*

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

5.5 Spatial interpolation of erroneous data

Extension of the dataset has included many satellite images that contain incorrect data. An example is shown in Figure 15, from 7 January 1990. 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 simplify error detection.

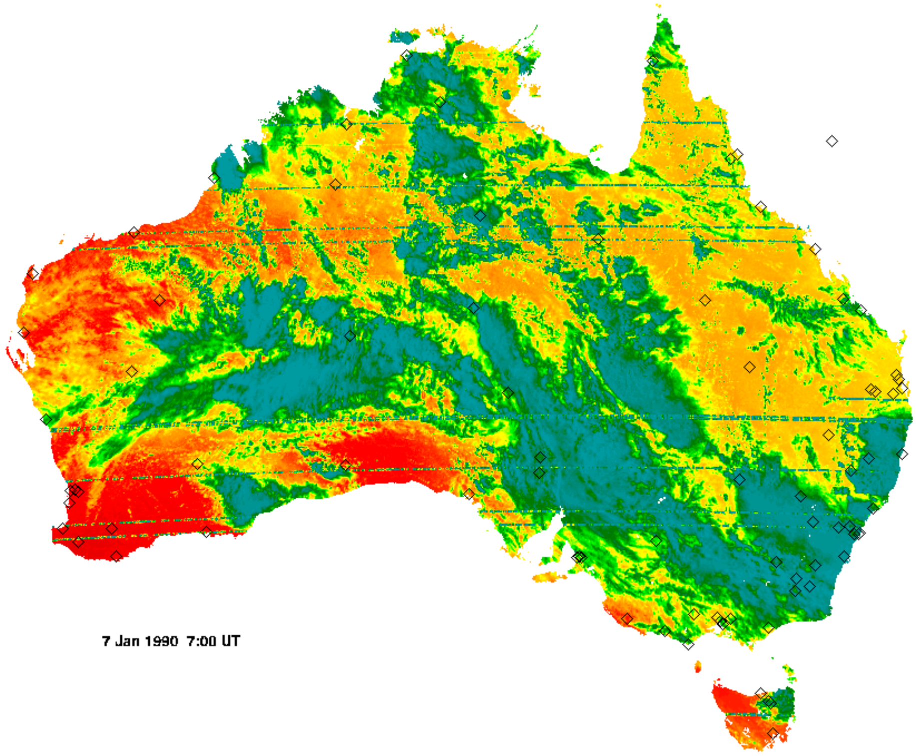


Figure 15. Satellite-derived direct normal irradiance for 7 January 1990, 07:00 UT.

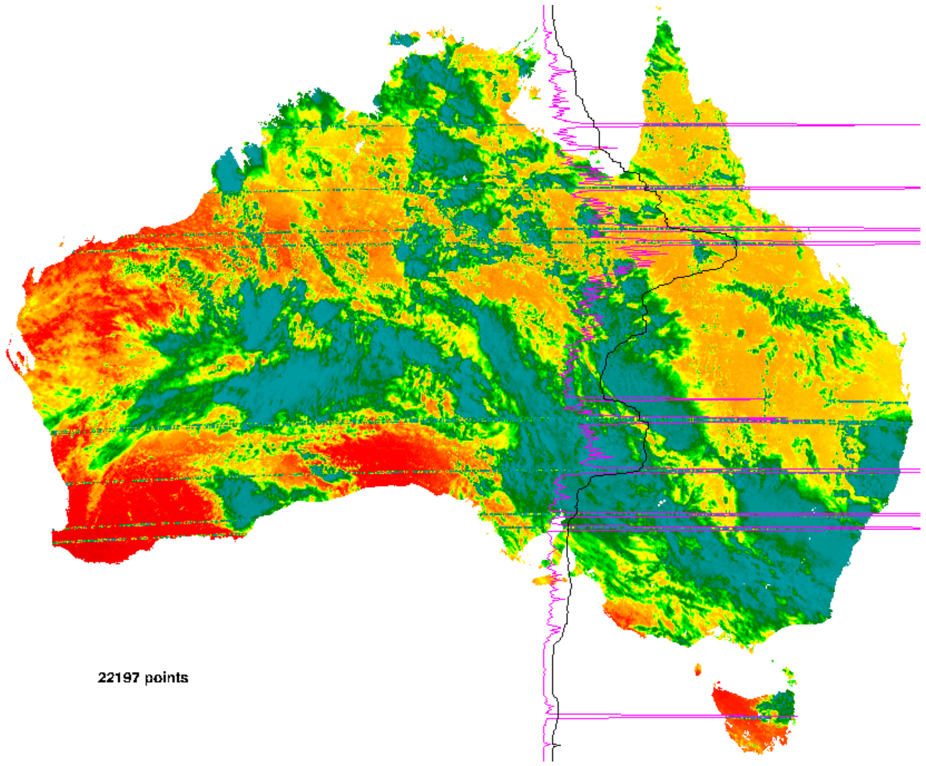


Figure 16. Satellite-derived direct normal irradiance as in Figure 15, showing errors detected by the algorithm described herein, and the number of pixels potentially affected.

Figure 16 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. The result, for this image, is shown in Figure 17.

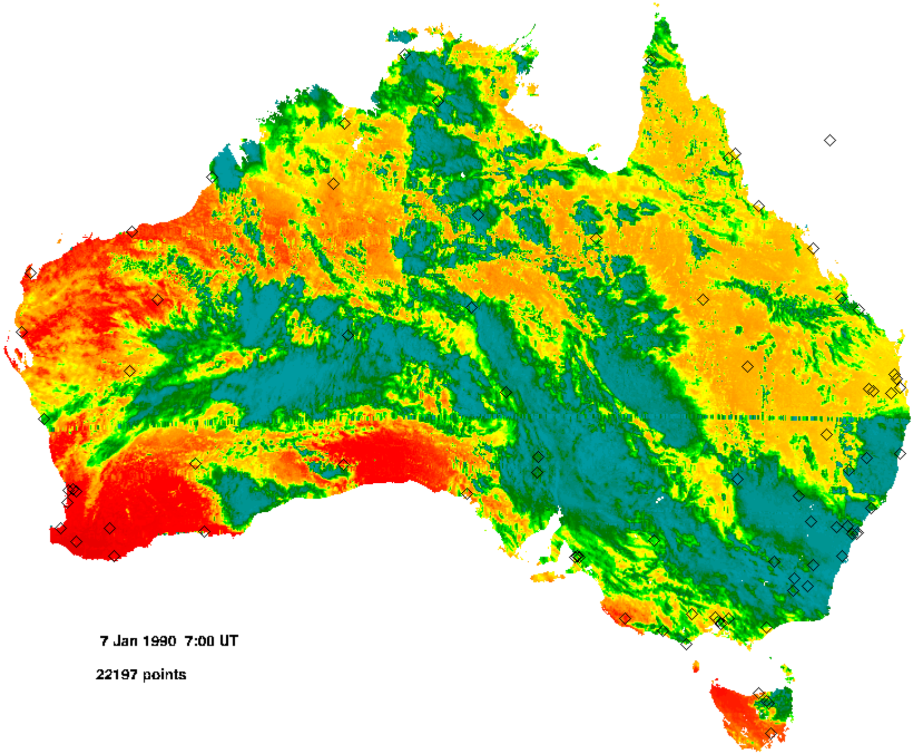


Figure 17. Satellite-derived direct normal irradiance as in Figure 15, after correction of errors detected as in Figure 16. Diamonds denote NatHERS representative sites.

Most seriously anomalous data are detected by this algorithm, but some visible anomalies miss detection or, as in Figure 17 around 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 may be distinctive.

5.6 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 & Le Marshall 2001). For example, the actual observation times in Figure 11 come from GMS-5 A data, so the times range from 03:47 to 03:53 UT, North to South. For the subsequent file labelled 0400 UT, from GMS-5 B data, observation times range from 04:40 to 04:46 UT.

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

Latitude	GMS-4	GMS-4	GMS-4	GMS-5	GOES-9	MTSAT-1R	MTSAT-2
Start date	1990/01/01	1993/01/01	1994/07/01	1995/06/11	2003/05/21	2005/11/01	2010/07/01
End date	1992/12/31	1994/06/30	1995/06/10	2003/05/20	2005/10/31	2010/06/30	ongoing
-10.0	45.7	47.2	46.7	46.7	39.9	46.2	44.7
-15.0	47.7	48.2	47.7	47.7	41.0	47.2	45.7
-20.0	47.7	49.3	48.8	48.8	42.0	48.3	46.8
-25.0	48.7	50.2	49.7	49.7	43.0	49.2	47.7
-30.0	49.6	51.1	50.6	50.6	43.9	50.1	48.6
-35.0	50.5	52.0	51.5	51.5	44.7	51.0	49.5
-40.0	51.2	52.7	52.2	52.2	45.5	51.7	50.2
-44.0	51.8	53.3	52.8	52.8	46.0	52.3	50.8
B shift	-7.0	-6.5	-6.2	-7.0	-12.0	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 of Meteorology with the initial release, the data have been checked against 1-minute measurements from Bureau instruments; those sites that measure *G*, *F*, and *R*, as used for hourly totals in NatHERS. The description includes:

- *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^{-2} and typically around +20 W m^{-2}. This is +4% of the mean irradiance of around 480 W m-2. The root mean square difference, calculated on a similar basis, is around 130 W m^{-2}, 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 18. As expected, the correlation was closest when the satellite-derived values were interpolated from measurement minute to the centre of the hour; it was then comparable to the figures quoted above.

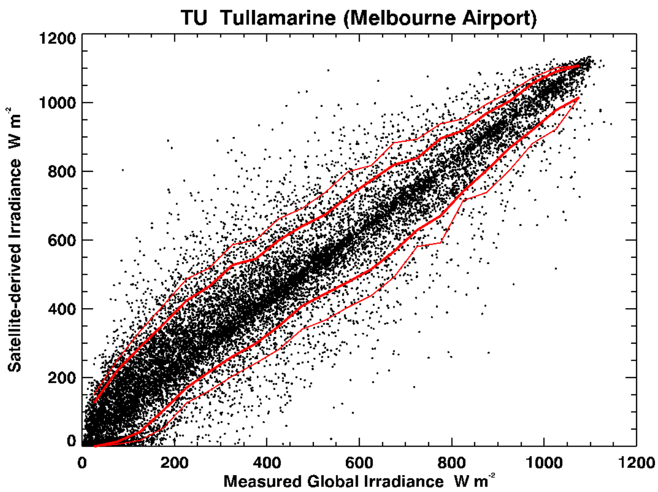


Figure 18. Measured hourly and satellite-derived instantaneous global irradiance interpolated to the same times at Tullamarine. Contours in red show the 5th, 10th, 90th and 95th percentiles.

As a further test, we examined the satellite data for the relationship illustrated in Figure 12 and Figure 13. As Figure 19 shows, the satellite-derived data match the ground-based data in this respect, and they serve as a very good proxy for hourly data.

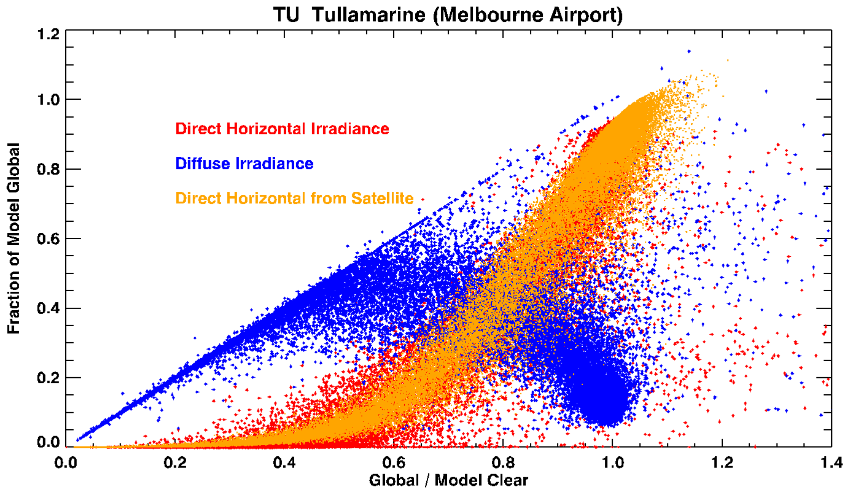


Figure 19. Diffuse (blue) and down welling direct (red) fractions of global radiation for measured data at Tullamarine, compared with satellite-derived values for the same location.

In other analyses (Liley 2011), we have found that the relationship between down-welling direct (or diffuse) and global irradiance, as a fraction of clear-sky global, is preserved in form as 1-minute data are aggregated into 10-minute or hourly values. This result is illustrated in Figure 20.

In 1-minute data, the vast majority of data fall in the sun-obscured ($G/G_c \lesssim 0.5$) or sun-out ($G/G_c \approx 1$) conditions, with few points in between except when there is very thin bright cloud. There are also many instances where the sun is out and scattered or broken white cloud is much brighter than blue sky, observable as 'cloud enhancement' with $G/G_c > 1$. In hourly totals for the same cloud cover, the cloud-enhanced values are mixed with sun-obscured so that the result falls along almost the same locus as for 1-minute values. There are however very few instances of cloud enhancement in hourly data as the the sun is rarely unobscured for the whole hour.

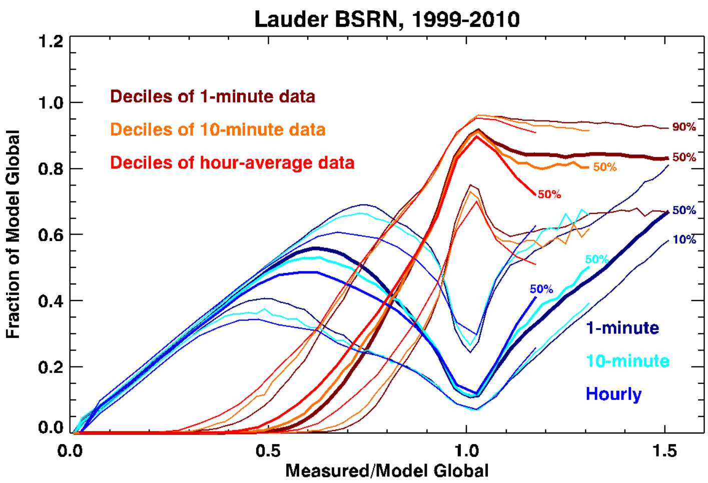


Figure 20. Relationships between diffuse and down-welling direct components of global radiation, as a fraction of clear-sky model values, are similar with time aggregation up to hourly.

It is notable in Figure 19 that the instantaneous values derived from satellite follow more closely the hourly pattern of Figure 20, with more frequent conditions that would correspond to partial obscuration of the sun, and minimal cloud enhancement. It is likely that this is a consequence of the spatial averaging in the satellite data. In conditions where there is scattered cloud, averaging over a cell approximately 5 km on a side has a similar effect to averaging over one hour at a single location.

6 Derivation of RMY/TMY files

6.1 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 21. 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 21 (i.e., in kWh m^{-2} day^{-1}) would require normalisation by standard deviation, interquartile range, or similar measures of dispersion, but they are sensitive to variation in the statistical distribution.

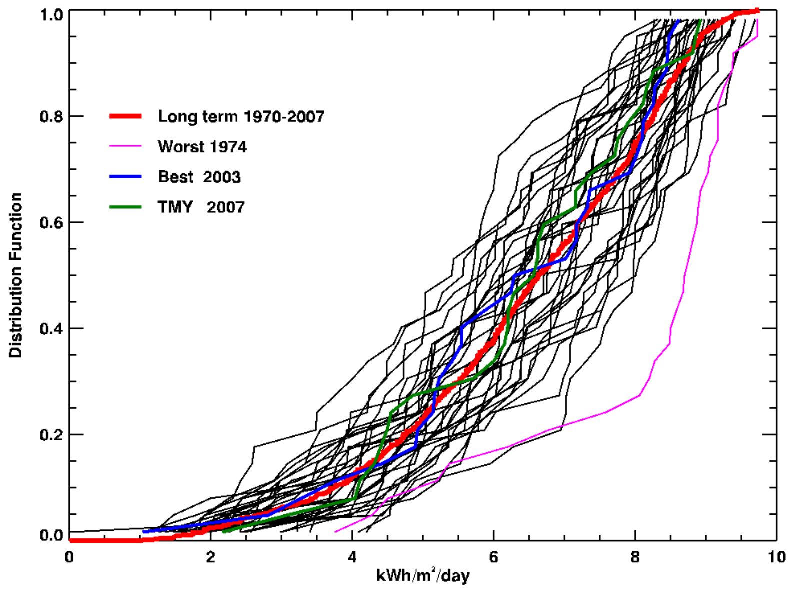


Figure 21. Distribution functions of January daily global irradiance for Auckland, New Zealand. The best match (2003) to long-term distribution is for irradiance only, whereas the TMY (2007) is chosen on the weighted sum of several parameters, and other considerations as described in the text.

The basis for selecting representative months for TMYs or RMYs then rests with the weighting given to different variables, usually including dry bulb and dew point temperatures (max, min, and mean), wind speed (max and mean), global and direct radiation (daily totals). In the 2008 revision of ACDB for NatHERS, three different weightings were used, as given in Table 4. The B and C versions were considered more relevant to larger buildings, as they give less weight to solar radiation.

Table 4. **Weightings for Finkelstein-Schafer statistics in NatHERS RMYs, 2008-2012.**

Index	RMYA	RMYB	RMYC
Max Dry Bulb Temperature	1	1	1
Min Dry Bulb Temperature	1	1	1
Mean Dry Bulb Temperature	2	2	2
Max Dew Point Temperature	1	1	1
Min Dew Point Temperature	1	1	1
Mean Dew Point Temperature	2	2	2
Max Wind Speed	1	1	1
Mean Wind Speed	1	1	1
Global Radiation	5	5	2
Direct Radiation	5		
Total (denominator)	20	15	12

The present revision has discontinued the RMYB and RMYC weightings, as it seems they are not much used, generate some confusion, and often differ little in the selection of representative months from the RMYA weightings. Correspondingly, the appended letter ‘A’ is dropped, and they are now just the NatHERS RMY files.

6.2 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 & 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 developed by NIWA and described below.

For the New Zealand TMYs (Liley 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.

7 Delineation of climate zones

7.1 Post Codes

Separation of the Australian continent into NatHERS zones is most readily achieved from an existing tessellation of the land mass, with certain requirements. It needs to have enough tiles to provide adequate resolution, with finer resolution in areas of high population density, and it should comprise boundaries of which some coincide with the geographical features that demarcate climatic regions.

Historically, this need has been met with Australian Post Codes, which have provided around 3000 polygons covering the continent. Their combined borders have included state boundaries, and they have generally enclosed comparable human population so that resolution is highest in densely populated areas. A disadvantage with post codes is that they are periodically updated to reflect population movement and, recently, a decline in postal volumes. Defining NatHERS zones in terms of post code boundaries requires those definitions to be revised whenever the post codes are revised, as happened with the NatHERS 2011 zones as shown by comparison of the zones in Figure 2 with those in Figure 1.

There are more stable alternatives to post codes, such as the Local Government Area (LGA) boundaries that appear in Figure 3. One convenience of post codes is that they are known or easily found for all house locations, with searchable indexes readily available on the internet. With interactive tools that work from a street address, and on-line maps as developed for NatHERS using the data behind Figure 2, it is no longer necessary to base NatHERS zones on post codes, LGA boundaries, or any other pre-existing tessellation. It may be better for future NatHERS updates to be independent.

A decision to that effect would change the way that any new NatHERS zone would be defined, and existing boundaries would similarly be redefined so as to depend only on State or other legislative boundaries and physical climate descriptors.

7.2 Maleny

In addition to the 69 existing NatHERS zones, a further 14 nationwide have been developed in the past or in the present work. The zones for Brisbane and Amberley are shown in Figure 22, together with the reference sites of potential zones.

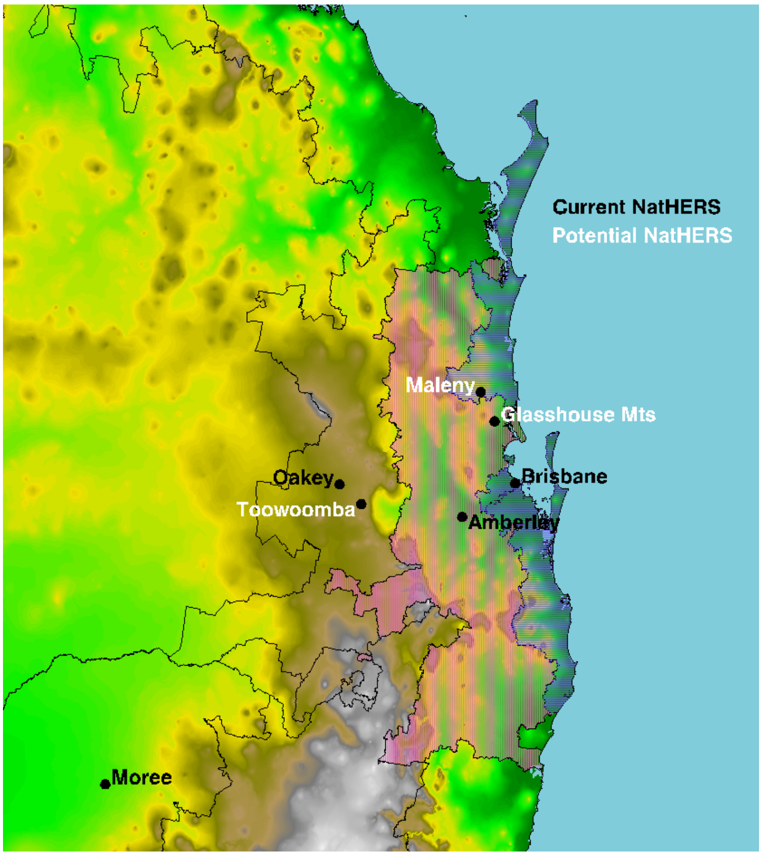


Figure 22. Boundaries of the Brisbane (BR, #10) and Amberley (AM, #9) NatHERS zones. Existing NatHERS sites are named in black, and potential new sites in white. Map colours show elevation.

The Brisbane zone (BR, #10) includes the coastal lowlands around the city and inshore islands. It extends south as far as Ballina and, separately, from the Sunshine Coast to Inskip and Fraser Island. Amberley extends from Washpool National Park in the south to the Gympie region in the north, and west into the foothills short of the high ground around Toowoomba. The new sites labelled in white prompt the question of which should be in included, and what region should be assigned to them.

Here we focus on this question for Maleny, represented as shown in Figure 23 by the daily climate station in the town at 425 m asl, but dependent on data from the Jimna Forestry AWS, somewhat higher (523 m asl) and further inland. Figure 23 also shows the climate stations in the area, both AWSs, which record hourly, and manually operated sites, typically recording once or twice daily.

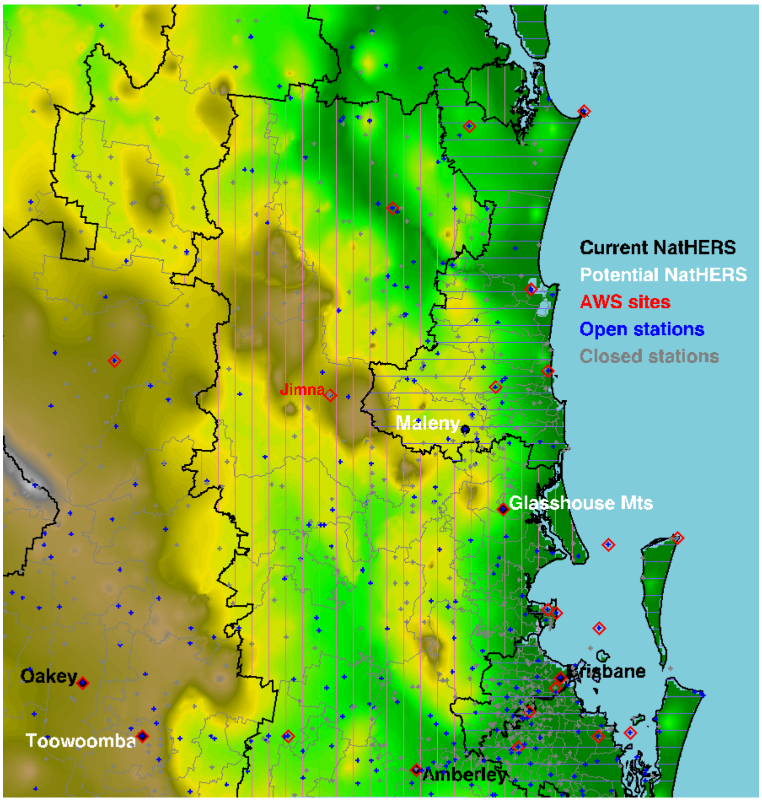


Figure 23. Present NatHERS (black) and post code (grey) boundaries in the vicinity of Maleny, Queensland. Also shown are the locations of Automatic Weather Stations (red diamonds) with hourly data, and other climate stations (+) that hold some climate records, though perhaps only rainfall. Many of those stations are closed (grey), but those marked in blue remain open.

When Maleny, and Toowoomba further west, were proposed in ACDB 2008, its authors undertook a very detailed study of the alignment of then post codes with Bureau of Meteorology climate maps. Ten post codes were recommended for inclusion in the Maleny zone; numbers 4514, 4519, 4521, 4550, 4552, 4570, 4571, 4574, 4600, and 4601. Their 2008 boundaries are shown in Figure 24. It is immediately apparent that the climate zone aligns fairly closely with the altitude, as represented by colour in the maps.

By analogy with the 2008 assignments, suggested current post code assignments are shown in Figure 25. They include post codes 4512, 4514, 4515, 4521, 4552, 4570, 4571, 4574, 4600, and 4601. From Bureau data, Maleny differs from Brisbane in the effects of altitude and distance from the coast, experiencing cooler nights and much higher rainfall (2000 mm per annum vs 1000 mm).

Post code 4515 might perhaps belong within the Toowoomba climate zone, along with 4601, as suggested from the ACDB 2008 work. Post code 4620, covering the former shire of Woocoo[1] (located inland of Maryborough), should perhaps also be in Maleny rather than in Oakey as at present.

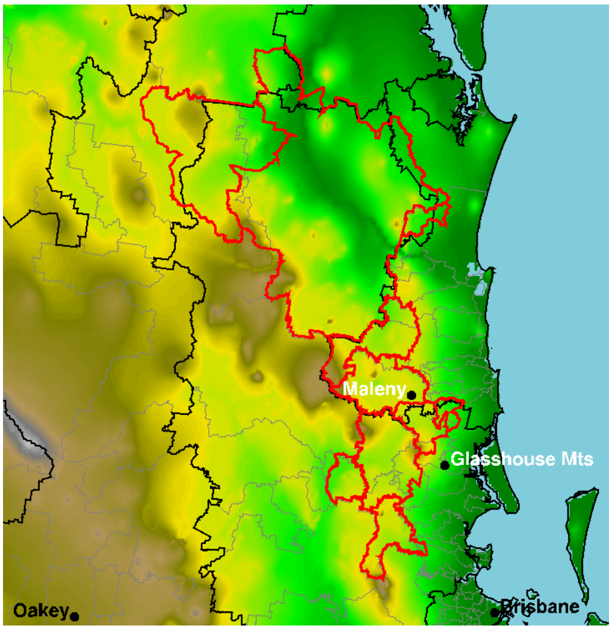


Figure 24. **As above, with post codes for inclusion in a Maleny zone as recommended in the 2008 revision of ACDB outlined in red. That revision introduced the Maleny zone, albeit based on data from the site reclassified here as Glasshouse Mountains.**

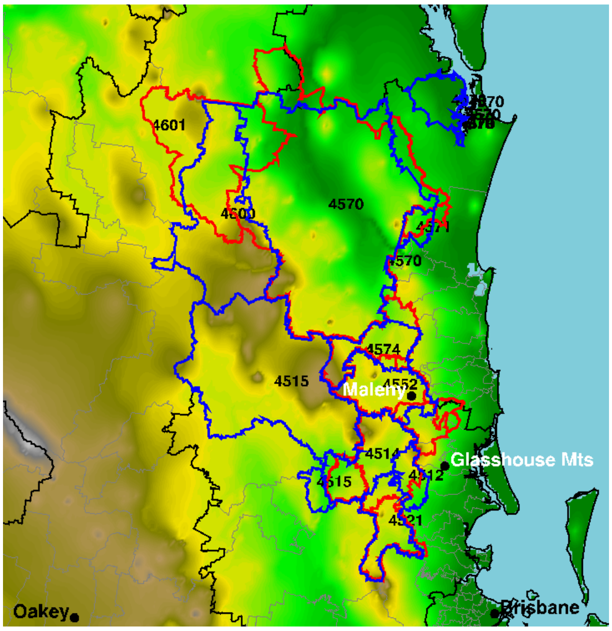


Figure 25. **By analogy to the 2008 recommendations, again in red, suggested current post code assignments are shown in blue.**

[1] The former shire of Woocoo was amalgamated in 2008 to form the larger 'Fraser Coast Regional Council' (along with areas of the former Maryborough City Council, Hervey Bay City Council and parts of Tiaro Shire Council. This change was part of statewide local government amalgamations at that time.

Figure 25 also illustrates the hazard of depending on post codes; some of the same codes are no longer applicable because boundaries have shifted or expanded and now include climatically dissimilar areas. This occurs for post code 4570 in the defined boundary for Amberley in NatHERS 2012, and again here for the putative Maleny boundary. As well as the Gympie region, it includes postal areas around Cooloola Cove and Tin Can Bay at sea level.

The Gympie region, 25 to 75 km inland and with higher average elevation, has the cooler nights of Maleny, but similar rainfall to Brisbane. The difference highlights a difficulty of making these assignments. While it is possible to use climate layers in a Geographical Information System or similar software in order to make informed choices, that does not accurately reflect the relevance to building energy simulation. The selection of typical months for RMYs follows a convention that building performance depends on climate variables in proportion to the weightings used with F-S statistics. It seems logical to use the same weights in climate zone delineation.

We have completed a basic analysis of this type. Available post-1990 data from amongst the climate stations shown in Figure 23 are used to develop approximate time series of the relevant data variables; temperature, dew point, wind speed, global and direct radiation. Daily climate indices as listed in the first column of Table 4 then define the long-term Distribution Function as illustrated in Figure 21. In our new method, the RMY file of each neighbouring NatHERS zone is used to generate daily summary values of the ten variables, for which we compute F-S differences for each month and sum them to give annual difference for each daily data type. They are then combined with the RMYA weights of Table 4 to give an overall difference.

The prescription as described was speculative, with many potential fish-hooks. The analysis is readily possible with hourly data from AWS sites, but the synoptic data from smaller climate stations are infrequent. Two measurements of temperature per day (usually at 09:00 and 15:00) give an indication of the diurnal temperature range, but they cannot confidently be compared with daily maxima and minima from hourly RMY values. In our analysis, wherever sites recorded only once or twice per day, comparator daily values for each RMY were computed just from those same times of day. This was necessary only for temperature, dew point, and wind, because hourly satellite-derived radiation data are available for all mainland sites.

Even allowing consideration of all post-1990 synoptic data with only one or two values per day of the required variables gives a limited selection, as listed in Table 5.

Table 5. Synoptic climate stations in the vicinity of Maleny, and nearest NatHERS match.

BoM	Name	Years	Latitude	Longitude	Alt.	NH
40040	Caloundra Signal Station	- 1992	-26.8017	153.1500	46.0	BR
40062	Crohamhurst	- 2004	-26.8094	152.8700	200.0	MN
40093	Gympie	1990 -	-26.1831	152.6414	64.5	AM
40100	Imbil Forestry	- 2011	-26.4619	152.6644	127.0	AM
40264	Tewantin Post Office	- 1996	-26.3919	153.0408	8.3	BR
40282	Nambour DPI	- 2007	-26.6431	152.9392	32.5	BR
40284	Beerburrum Forest Station	1990 -	-26.9586	152.9619	48.0	BR
40318	Kirkleagh	- 1991	-27.0258	152.5642	103.6	MN
40651	Jimna Forestry	1990 -	-26.6644	152.4606	523.0	MN
40850	Baroon Pocket Dam	1992 -	-26.7150	152.8719	248.0	MN
40861	Sunshine Coast Airport	1994 -	-26.6006	153.0903	3.3	BR
40908	Tewantin RSL Park	1995 -	-26.3911	153.0403	6.4	BR
40988	Nambour Daff - Hillside	2007 -	-26.6442	152.9383	53.2	MN

Results are illustrated in Figure 26, and they are encouraging. Similarity was calculated for the six nearest NatHERS zones: MN, #73, Maleny; TW, #70, Toowoomba; OA, #50, Oakey; AM, #9, Amberley; BR, #10, Brisbane; and GL, #36, Gladstone. As shown in the last column of Table 5 and in Figure 26, only three of the six zones best represented any of the 13 climate stations. Station 400284 would obviously have gone with Glasshouse Mountains if that site were included in NatHERS, but otherwise it is more like Brisbane than Maleny, as noted previously.

As expected, the coastal stations are more like Brisbane, and the higher inland sites more like Maleny. Nambour DPI (pre-2007) and Nambour Daff – Hillside (post-2007) seem to be on the cusp, with the former assigned to Brisbane and the latter, slightly higher, site to Maleny. As they are only 150 m apart, the difference is probably spurious, or could result from the different time periods represented.

The analysis does answer the earlier question about post code 4570, enclosing Gympie and Imbil Forestry sites. Though the difference in F-S values (not shown) is not large, those stations are better represented by the Amberley NatHERS file. The post code list for Maleny is then just 4512, 4514, 4515, 4521, 4552, 4571, 4574, 4600, and 4601.

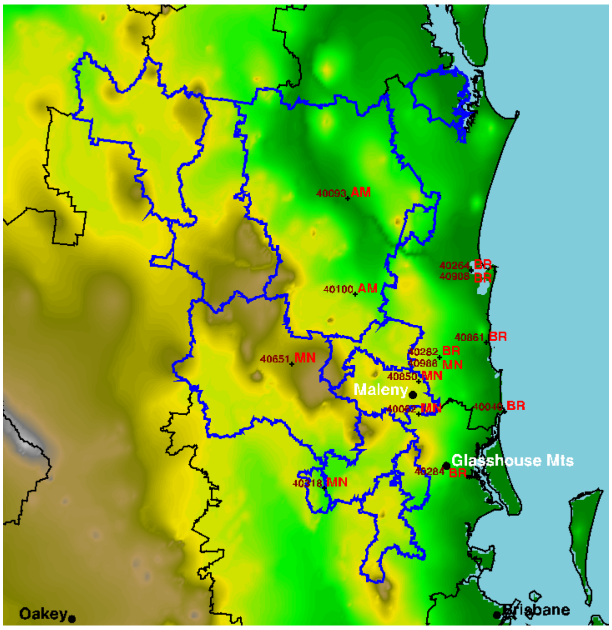


Figure 26. **Results from calculating the most similar NatHERS zone to 13 synoptic climate stations using weighted F-S statistics. Postcode 4570, around Gympie and Imbil Forestry, is better represented by Amberley than by Maleny, according to this analysis.**

Though this technique looks very promising, the spatial density of open or recent climate stations is too low to represent all post codes. Even if a system of polygons other than post codes is chosen as the basis for NatHERS zones, the same issue arises. To apply the technique widely and consistently, we need a suitable high-resolution representation of daily extremes in temperature, humidity, and wind, alongside the now excellent data for solar radiation.

For this purpose, a likely candidate is the Queensland Government's SILO Patched Point Dataset, which interpolates from open stations to provide time series for any of over 4600 climate stations across the continent. Gridded data from their interpolation might be even better, as the resolution matches the 0.05° (~ 5 km) grid of the radiation data. Unfortunately, funded time and delivery dates on current work did not allow adequate assessment of this alternative, but it should be considered in future work.

7.3 Willis Island and Christmas Island

Because the satellite radiation data do not include Willis Island, no update of it was possible, and we do not have any information on the reliability of the 1977-1995 global irradiance data. The Bureau of Meteorology has maintained a BSRN station on Cocos (Keeling) Island in the Indian Ocean, with much the same climate as the more populous Christmas Island, so we have created a new zone with data from the former but named for the latter.

The plots in Appendix A show that Christmas Island has even less variation in temperature than the small amount at Willis Island, and it has comparable values and variation in dew point and wind speed. It would be useful to use both files in the same building simulation to see whether differences are discernible. If not, the greater reliability of the radiation data for Christmas Island should favour its use for distant Australian tropical islands in the Indian or Pacific oceans.

7.4 Inshore islands

A further question arises with regard to the many small islands off the coast of Queensland and Western Australia, which are not covered in the satellite-derived radiation dataset. Close to the coast, the radiation data are extrapolated offshore, but at some distance the island should be better represented by Willis Island or the new Christmas Island.

Table 6. Synoptic island climate stations off the Pacific Coast of Queensland. Distance from the continent in kilometres is given in the Dist. column, and NatHERS match in NH.

BoM	Name	Years	Latitude	Longitude	Alt.	Dist.	NH
31037	Low Isles Lighthouse	1990 -	-16.3842	145.5592	2.9	16	RO
31192	Green Island	- 2010	-16.7614	145.9719	3.0	13	WS
32141	Lucinda Point	1990 -	-18.5203	146.3861	10.0	5	MK
33106	Hamilton Island Airport	2002 -	-20.3658	148.9536	58.7	12	MK
39059	Lady Elliot Island	1990 -	-24.1116	152.7161	3.6	78	RO
39122	Heron Island Res Stn	1990 -	-23.4417	151.9125	3.3	66	RO
39304	Heron Island	- 1998	-23.4483	151.9178	8.0	66	RO
200001	Middle Percy Island	1999 -	-21.6628	150.2711	208.7	58	WS
200701	Frederick Reef	1990 -	-20.9375	154.4019	4.9	414	WS
200732	Holmes Reef	1990 -	-16.4683	147.8734	1.9	212	WS
200736	Creal Reef	1990 -	-20.5303	150.3773	1.7	134	RO
200783	Flinders Reef	1990 -	-17.7195	148.4478	2.6	207	RO
200840	Bougainville Reef	1992 -	-15.4877	147.1183	0.0	188	WS

Applying the F-S technique above to this question gave results as shown in the last column of Table 6 and in Figure 27, based on similarity of the stations to NatHERS zones: WS, #31, Willis Island; CN, #32, Cairns; TO, #5, Townsville; MK, #35, Mackay; RO, #7, Rockhampton; and GL, #36, Gladstone.

Overall, the results are less convincing than those for Maleny, perhaps largely because in the absence of offshore radiation data the analysis is based only on temperature, humidity, and wind speed. Most of the islands further offshore are matched to Willis Island, but not all, with both Flinders and Creal Reefs found to be similar to Rockhampton. That Rockhampton was also matched by Low Isles Lighthouse north of Cairns suggests that more work on this topic is needed.

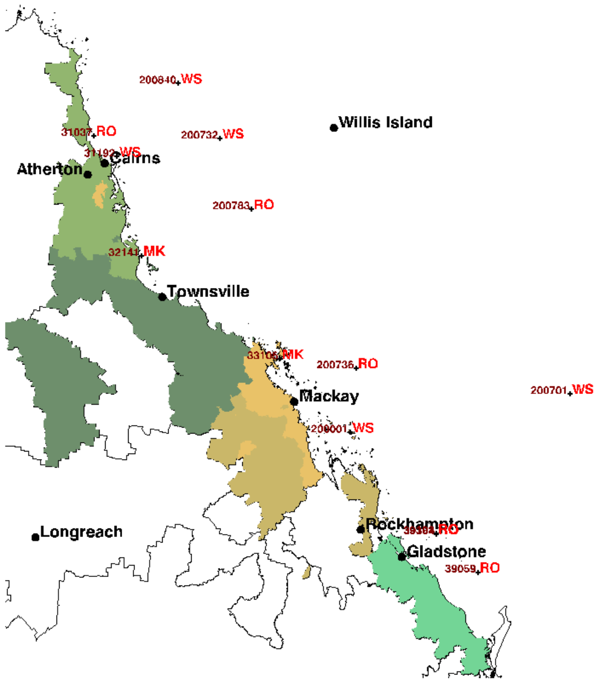


Figure 27. Results from calculating the most similar NatHERS zone to 13 island climate stations using F-S statistics. Analysis (not shown) using XI instead of WS gave the same results.

In the interim, it seems that mainland RMYs are applicable out to perhaps 100 km offshore, and beyond that the remote tropical island RMY should be used.

A final test was to replace the WS, #31, Willis Island RMY in the above analysis with the new XI, #83, Christmas Island RMY. That analysis gave the same results, matching XI wherever WS had been assigned. Thus, it seems reasonable to use the new Christmas Island RMY because of the higher quality of its solar radiation data.

8 Conclusions

Climate data files have been derived afresh from Bureau of Meteorology data, without recourse to past ACDB versions but for the same sites. The data are based on the period from 1990 to the present, coinciding with the satellite-derived radiation coverage and better ground-based instrumentation and data collection.

All data have been subjected to intensive quality assurance, with correction where possible. Major corrections have included spatially interpolated pressure data for those sites with no measurements, and marking anomalous values of all variables by a robust statistical procedure. A very widespread problem where the wet-bulb wick dries out, so that wet-bulb temperature tracks dry-bulb to very high temperatures, can lead to very unphysical values of implied humidity. Our previously-developed algorithms to detect and correct this problem have been further refined.

Marked changes over time in average wind speeds were found at many sites, and extra work was needed to resolve the cause and determine how best to treat the data. Review of the instrument history for the many affected sites showed that some wind data were at 8 m mast height and others nearer the surface, and anemometers are prone to other problems. We used a combination of data scaling between sites or instrument period, and simple exclusion of some data from RMY selection.

In many instances, anomalous data are simply flagged in the time series and thereby excluded from use in RMY file. The data flags for all variables should be considered in any use of the full time series.

Satellite-derived estimates of global and direct solar radiation data at high spatial resolution, acquired from the Bureau of Meteorology for 1990 to 2015 inclusive, have been used alongside high-quality ground-based data where available. The latter have been aggregated from 1-minute data to hourly totals in Mean Solar Time, to which the satellite-derived data have also been interpolated. Simultaneous collocated values from the two sources show excellent agreement.

The RMY derivation has used international best practice, extended with a more precise and objective method where there was some ambiguity in the standard references.

We have explored methods for objectively delineating NatHERS zones using the same criteria as in RMY selection, and applied the method for the new Maleny zone. Though it seems to work well, reliable climate data at higher resolution are needed. For this, the Queensland Government SILO datasets probably should be considered.

Coverage has been extended to 83 climate zones with the addition of Christmas Island as a valid replacement for Willis Island, applicable to distant tropical islands.

9 References

Chen, D. (2016). AccuRate and the Chenath engine for residential house energy rating.

Delsante, A. (2005). Description of weather data files used by AccuRate. Highett, VIC, Australia, CSIRO Manufacturing & Infrastructure Technology.

Energy Partners (2008). Australian Climate Data Bank - Weather data enhancement for Reference Meteorological Years. No. 86 p.

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: Jones, S.; Reinke, K. (eds). Innovations in Remote Sensing and Photogrammetry, pp. 161-172. Lecture Notes in Geoinformation and Cartography. Springer, Berlin.

Liley, J.B. (2010). Correction of Sydney Reference Meteorological Year data, and quality assurance for all NatHERS files. No. LAU2010-01-JBL. iii, 38 p.

Liley, J.B. (2011). Diffuse and direct components of solar radiation in Australia and New Zealand. Extreme Weather 2011. Wellington, Australian Meteorological and Oceanographc Society & NZ Meteorological Society.

Liley, J.B. (2013). Australian climate data and Reference Meteorological Years for NatHERS 2012. No. LAU2013-02-JBL. iii, 35 p.

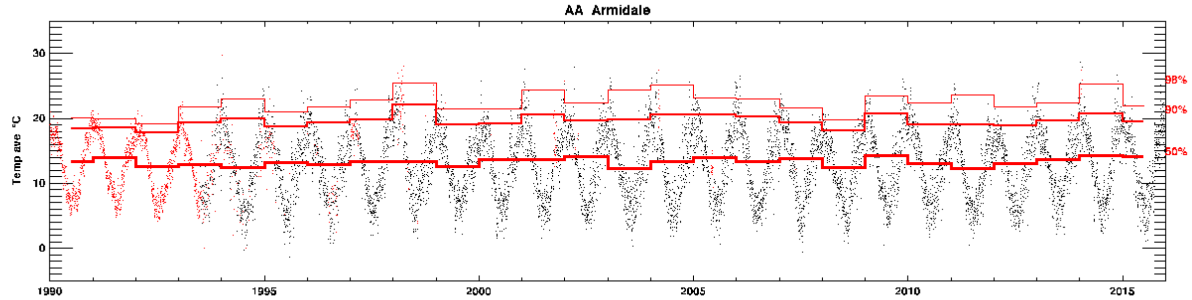
Liley, J.B.; Sturman, J.; Shiona, H.; Wratt, D.S. (2008). Typical Meteorological Years for the New Zealand Home Energy Rating Scheme. No. LAU2008-01-JBL. iii, 46 p.

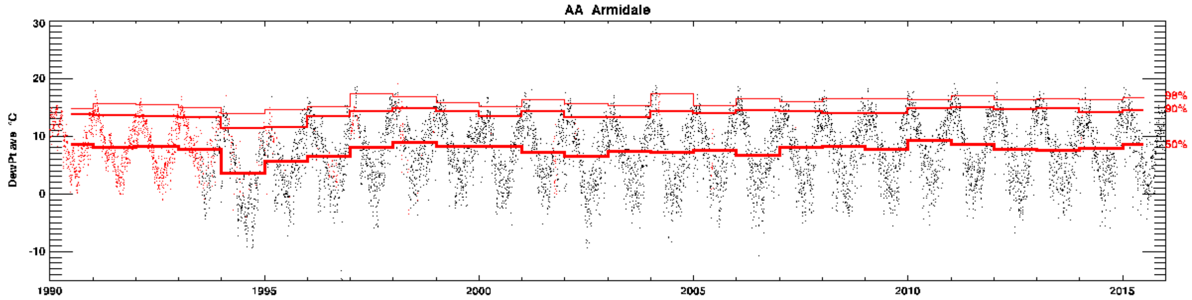
Marion, W.; Urban, K. (1995). User's Manual for TMY2s (Typical Meteorological Years). No. NREL/SP-463-7668. 55 p.

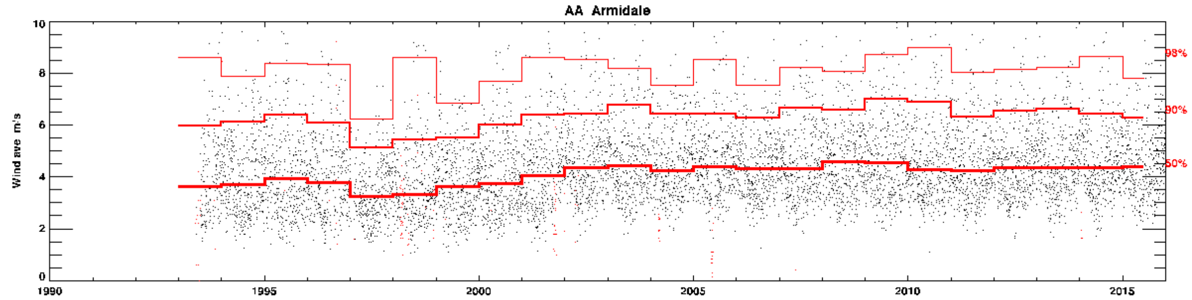
Weymouth, G.T.; Le Marshall, J.F. (2001). Estimate of daily surface solar exposure using GMS-5 stretched-VISSR observations. The system and basic results. Australian Meteorological Magazine 50: 263-278.

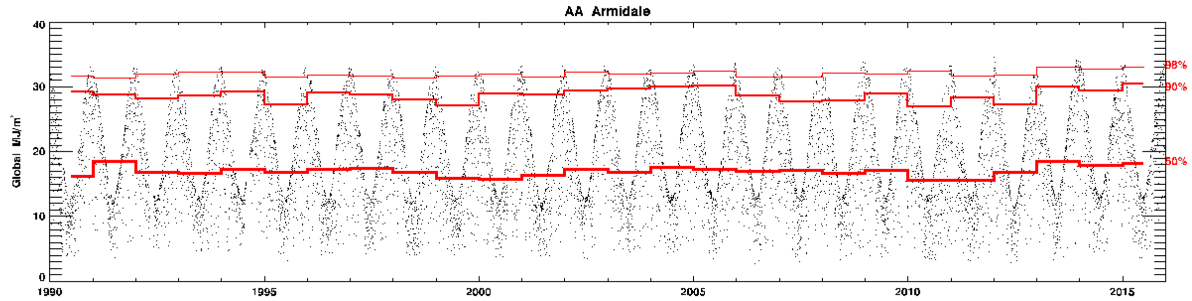
Appendix A Example time series

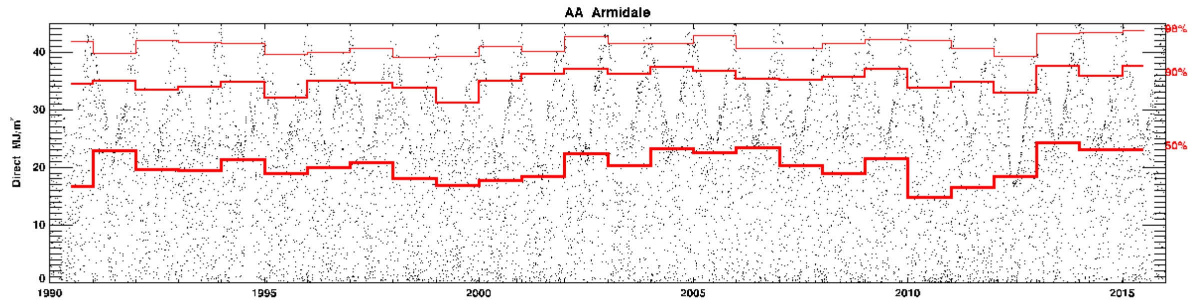
Example NatHERS time series for **Armidale**, showing temperature, dew point, wind speed, global and direct radiation, and annual percentiles (red lines) to assess consistency. Points in red are spatial interpolates, used only if required.



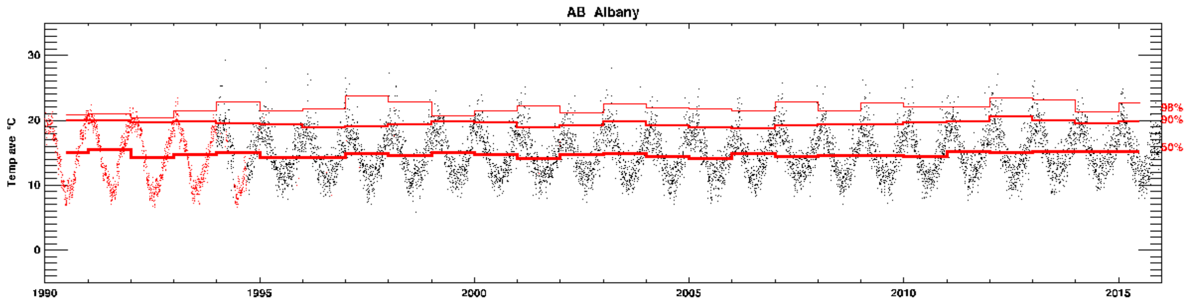


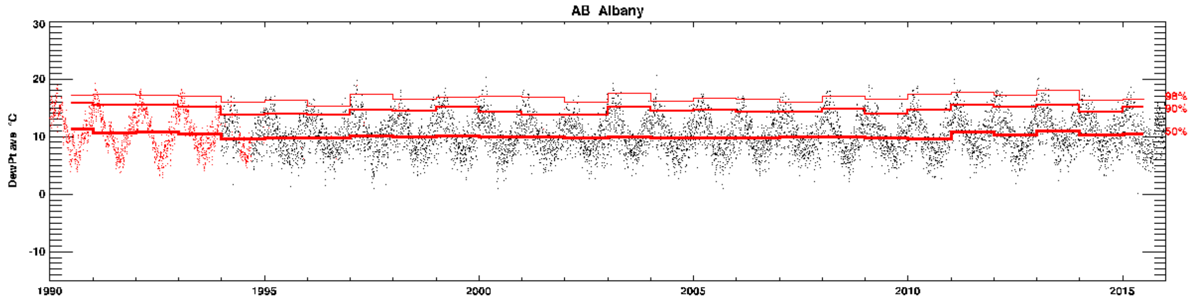


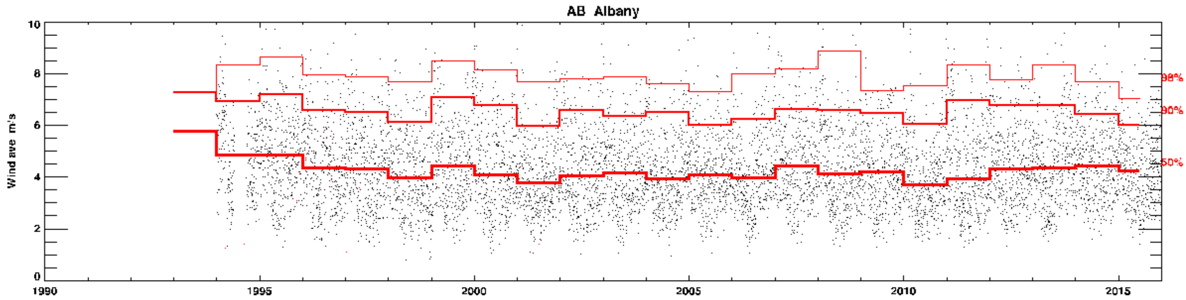


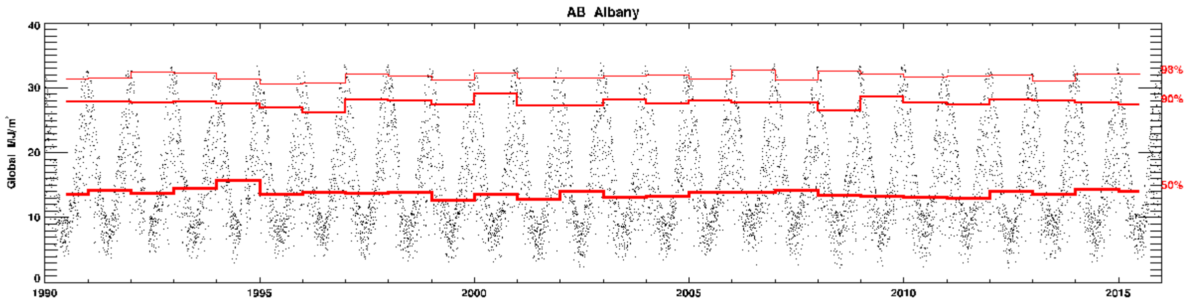


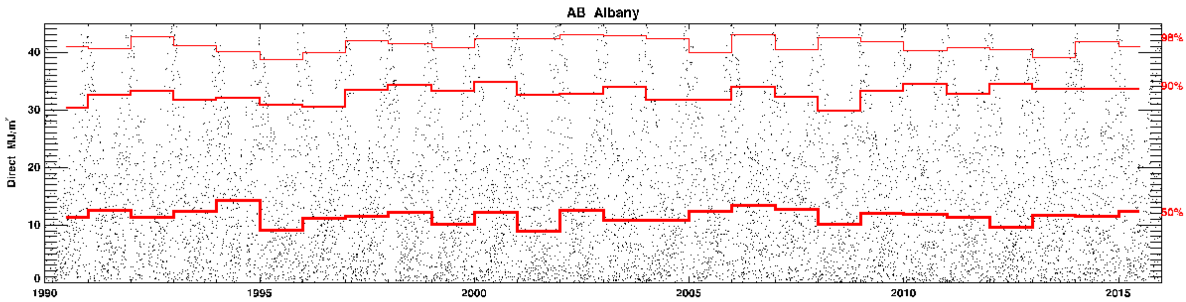
As above for **Albany**.



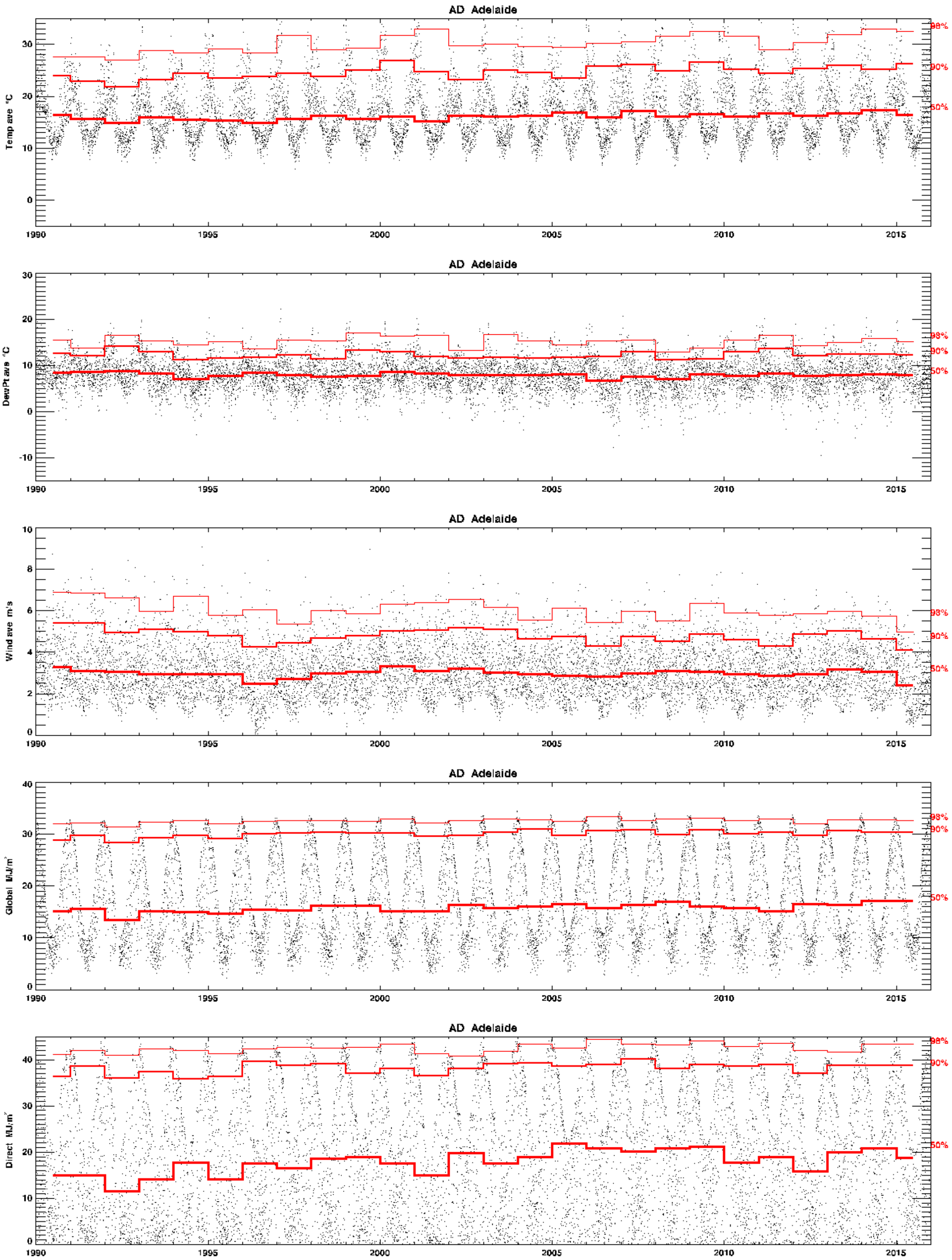




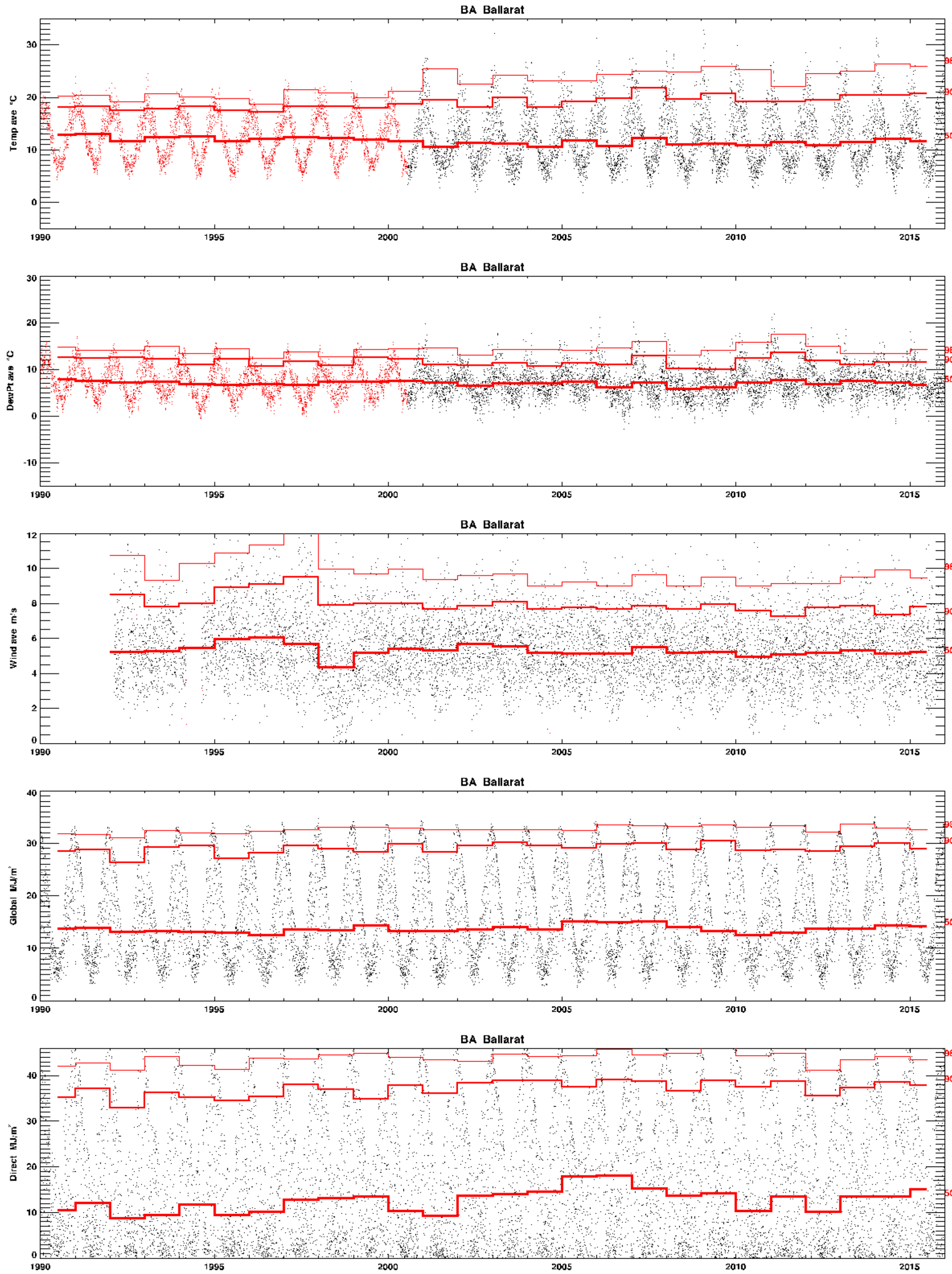




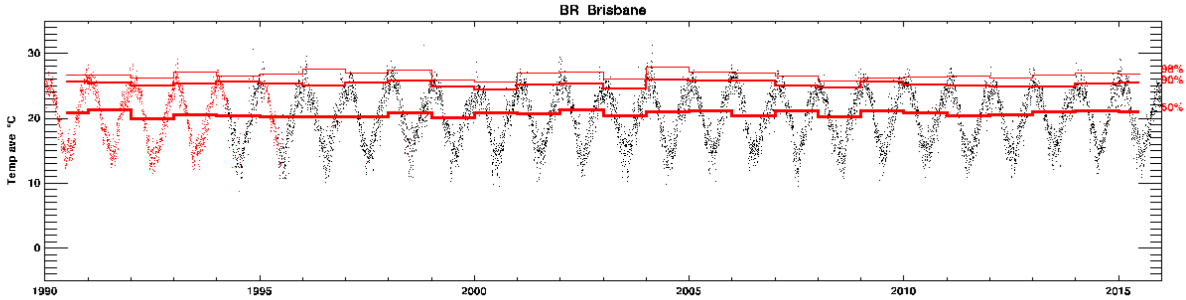
As above for **Adelaide**.

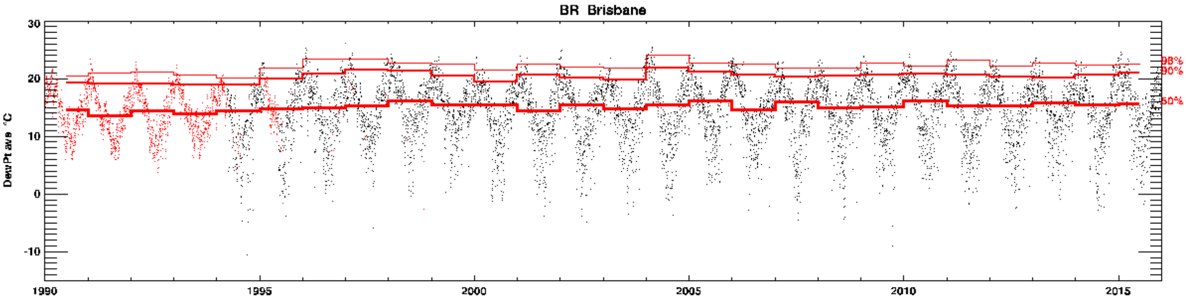


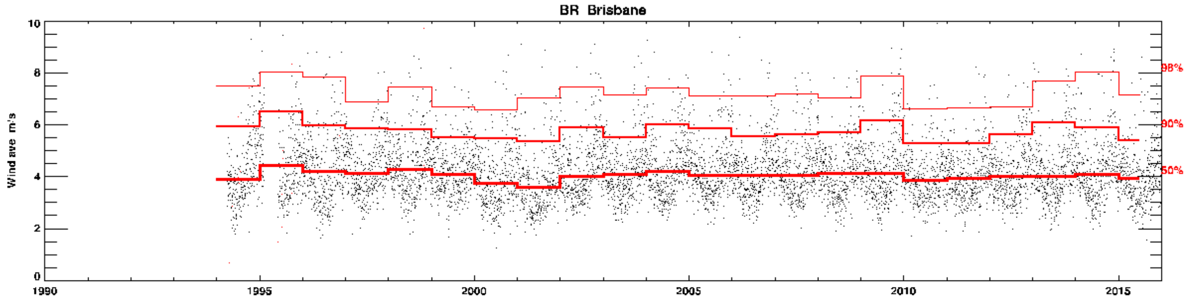
As above for **Ballarat**.

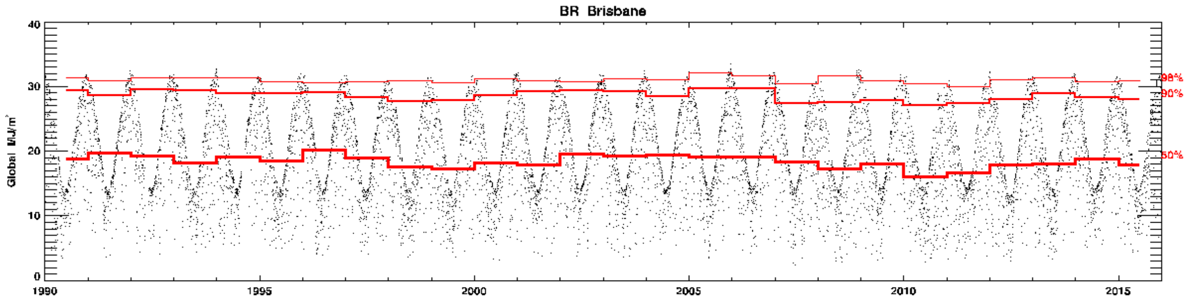


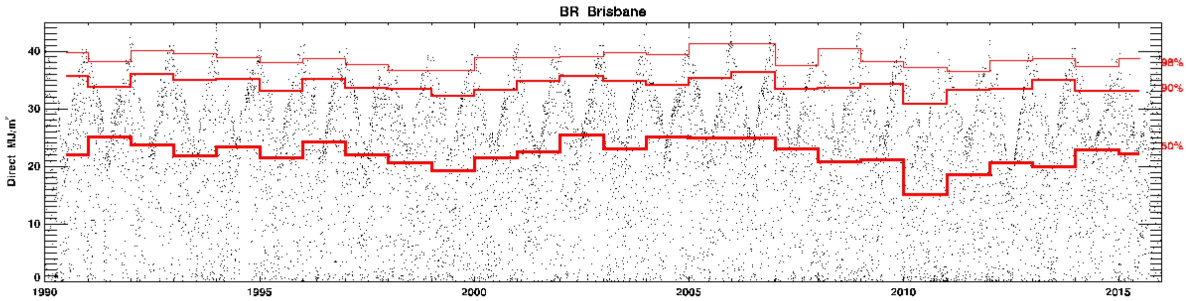
As above for **Brisbane**.



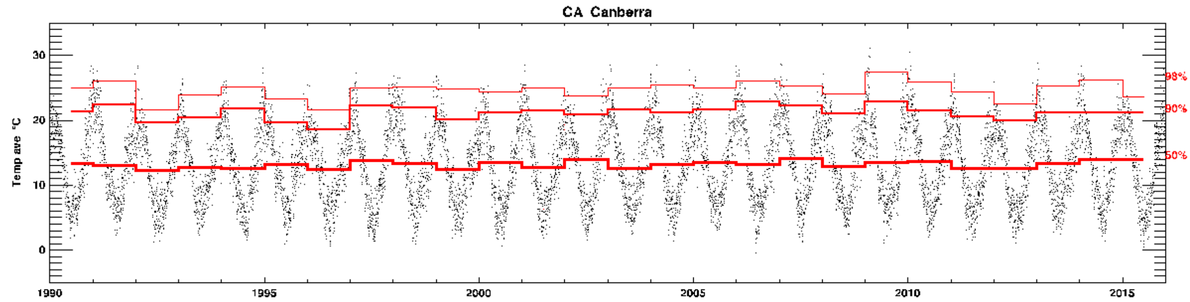


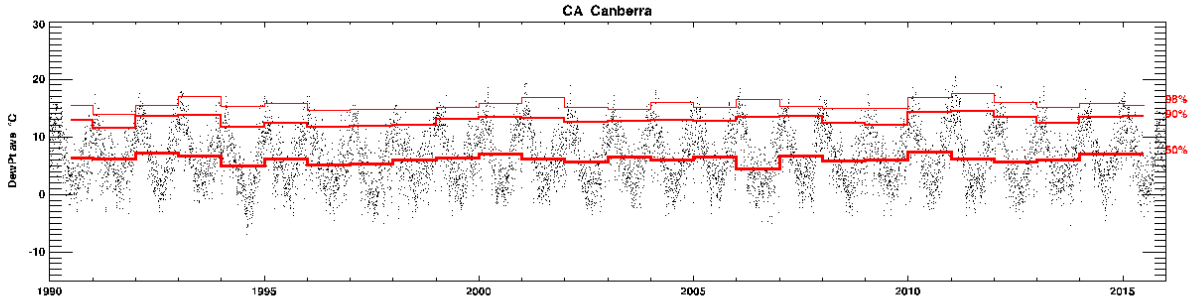


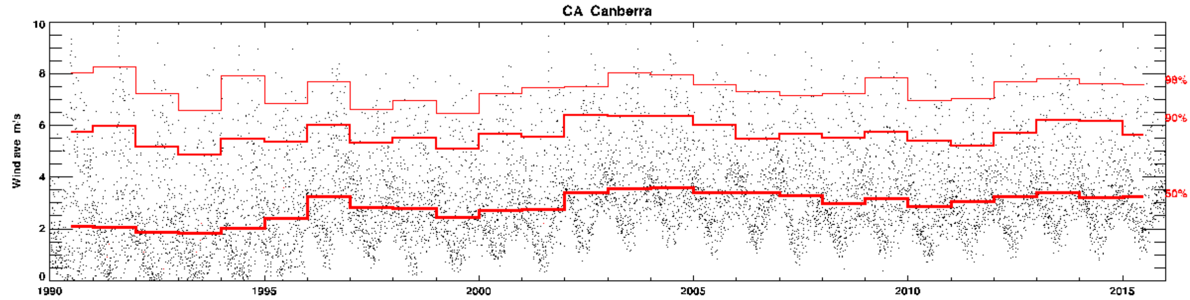


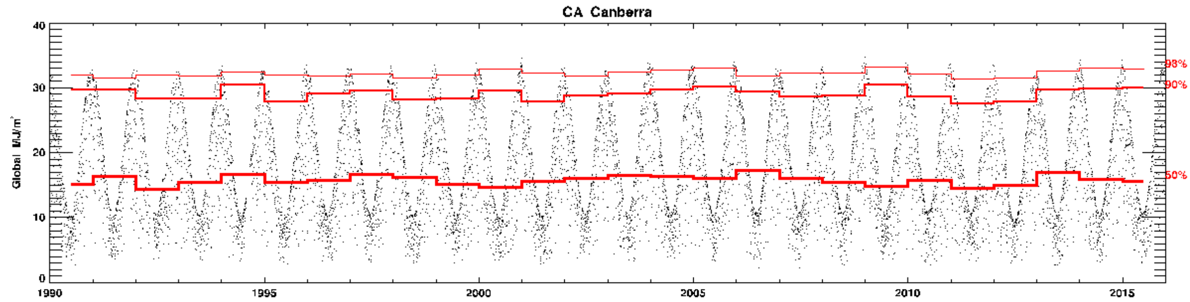


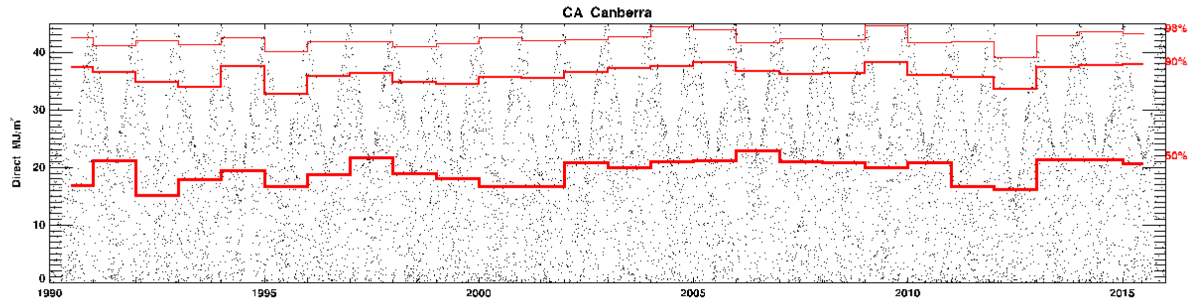
As above for **Canberra**.



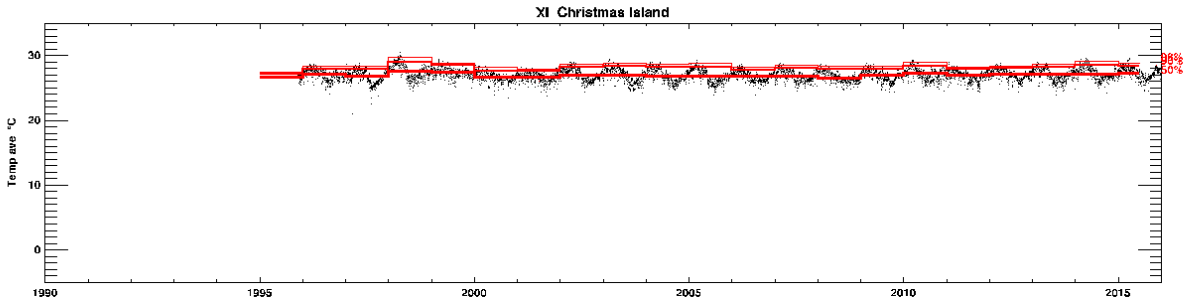


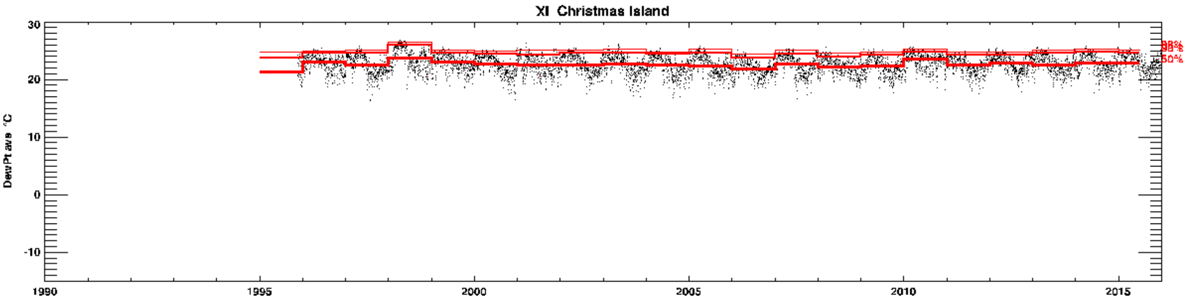


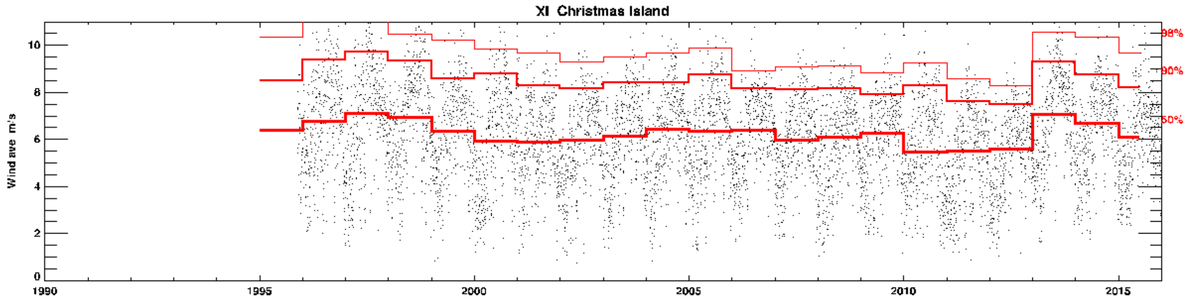


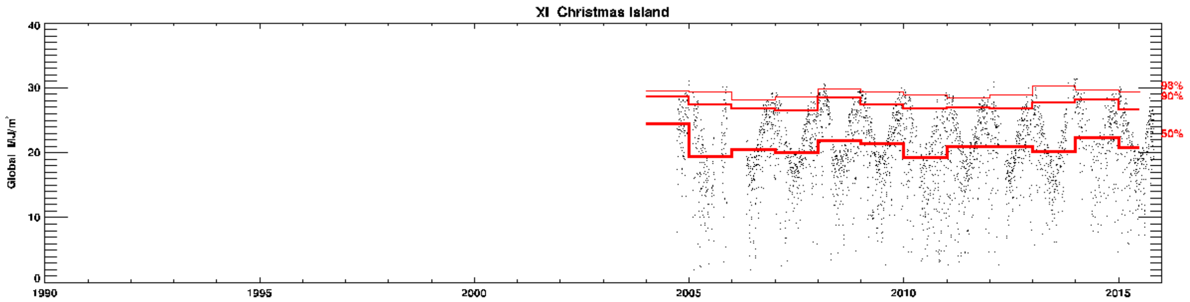


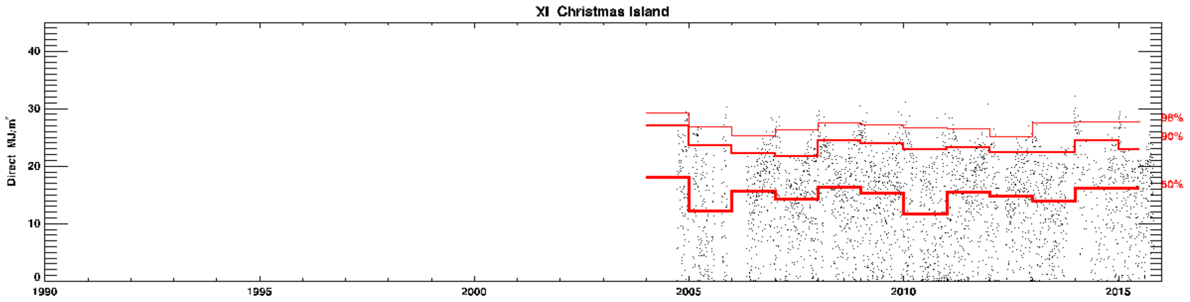
As above for **Christmas Island**.



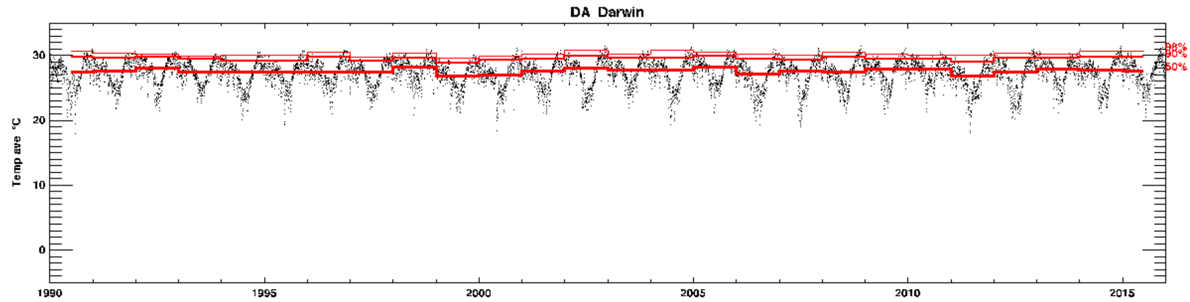


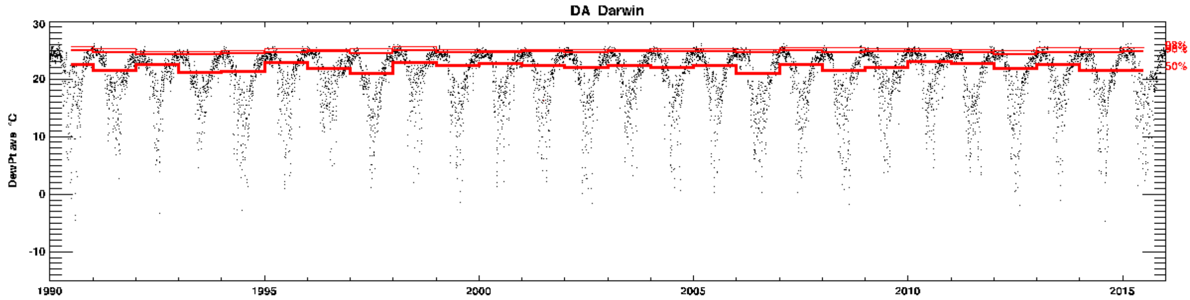


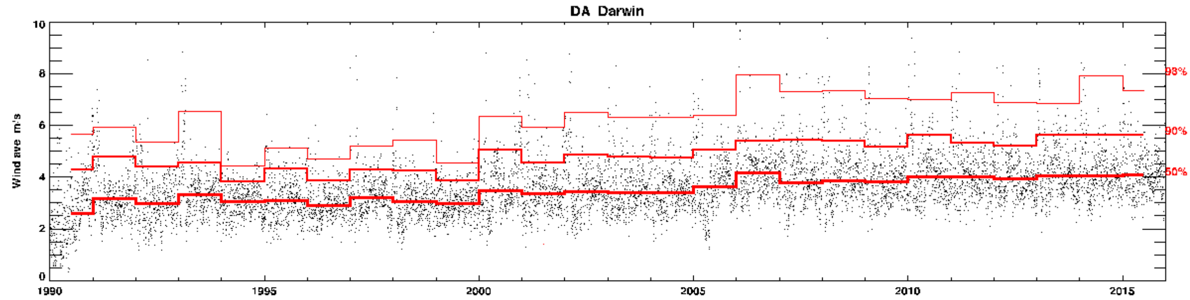


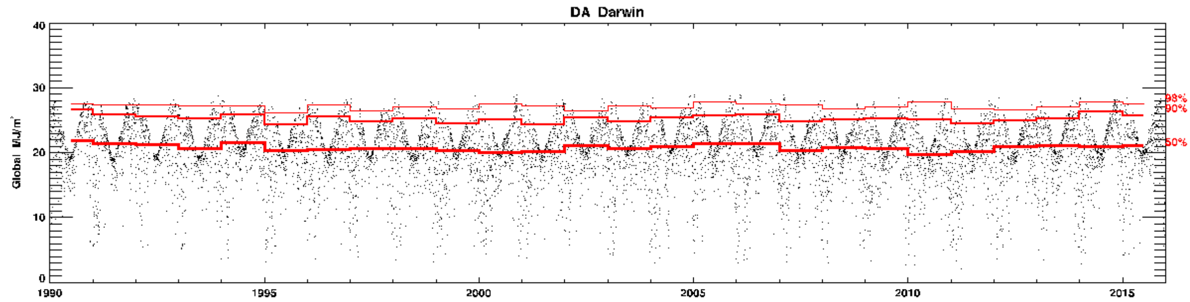


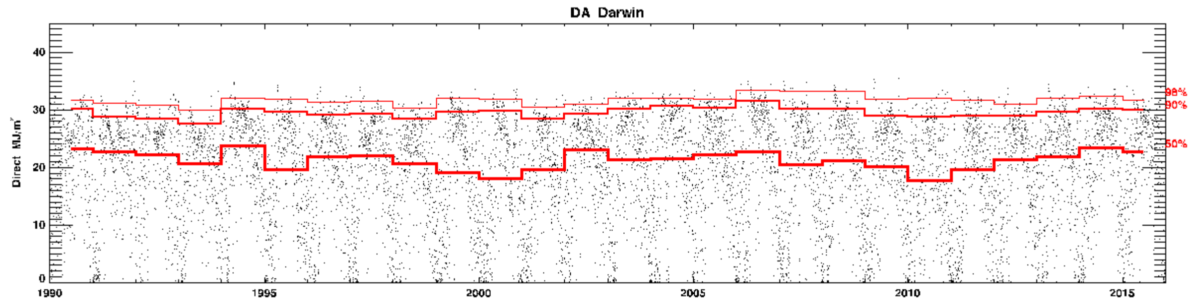
As above for **Darwin**.



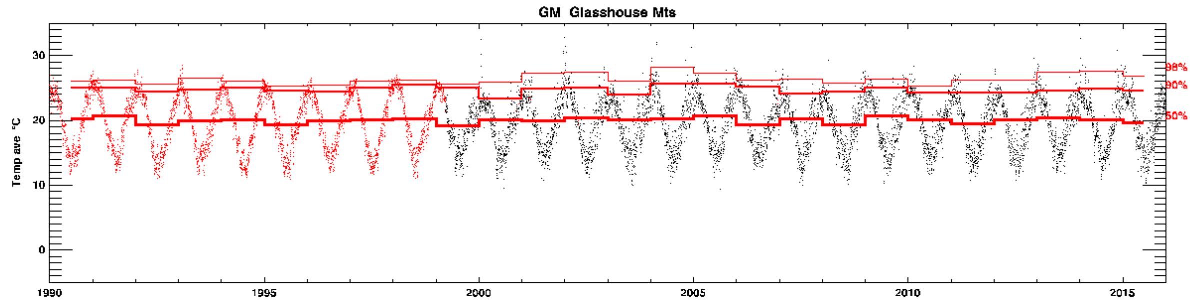


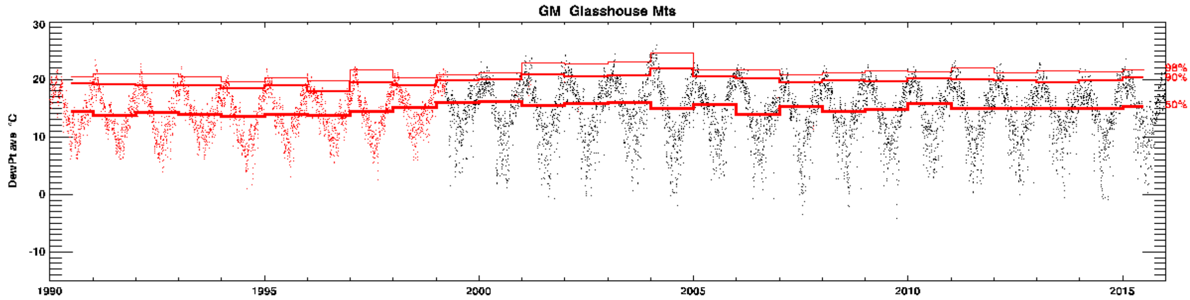


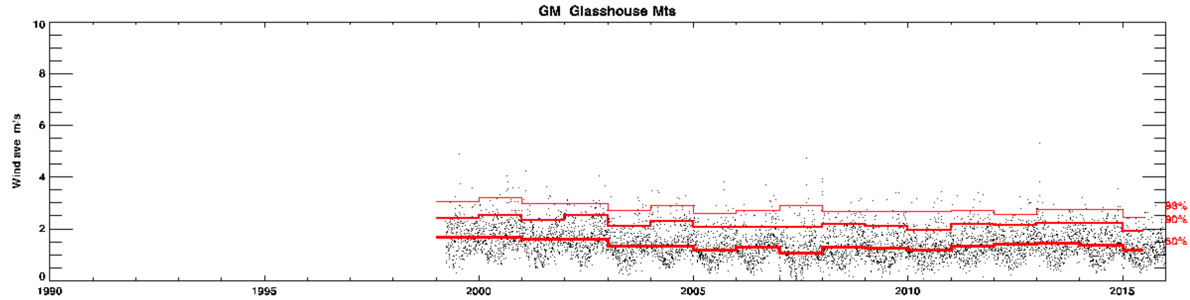


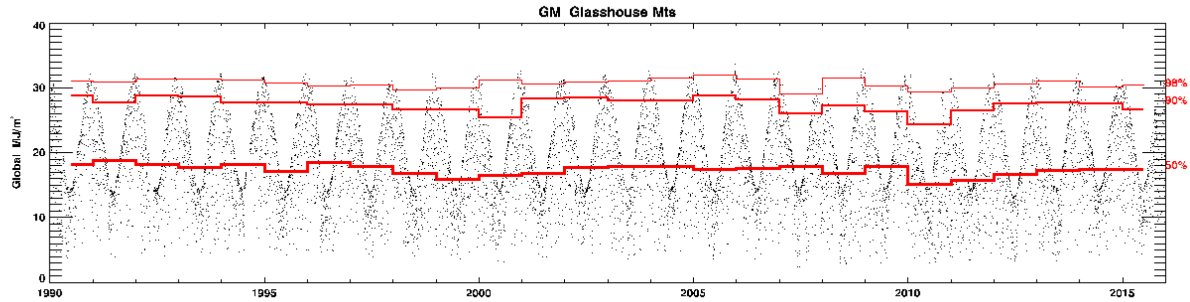


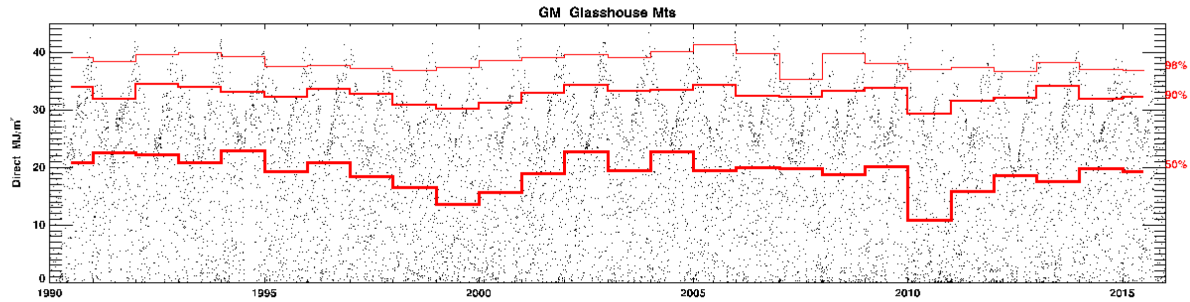
As above for **Glasshouse Mountains**.



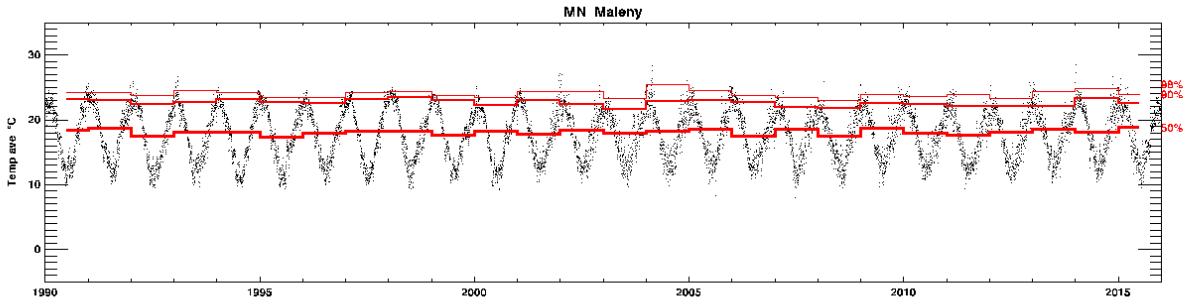


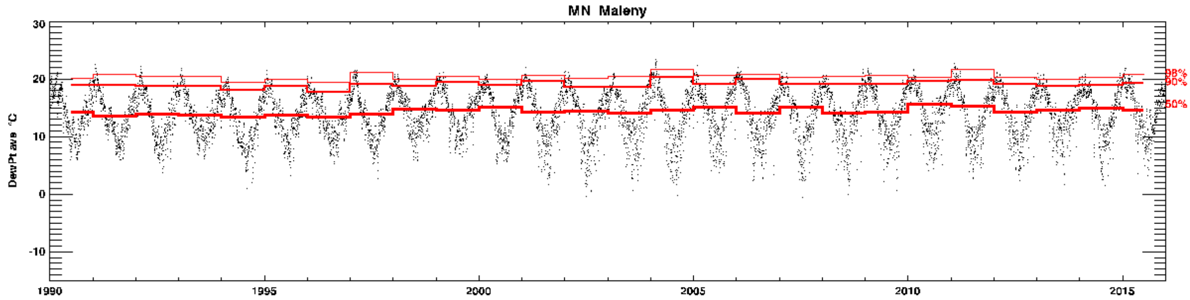


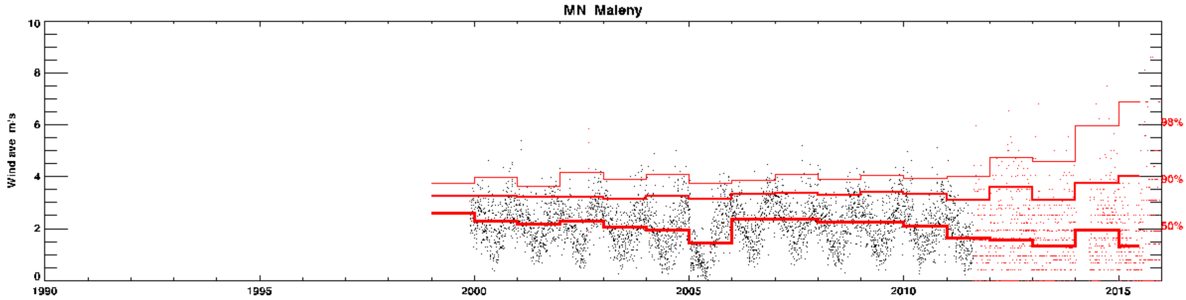


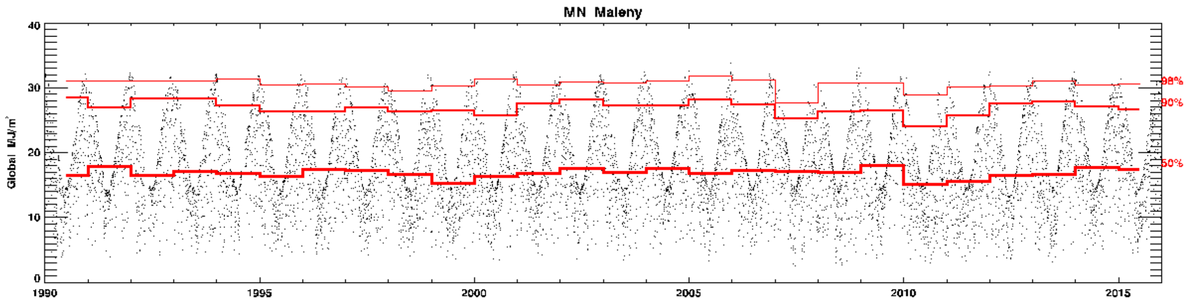


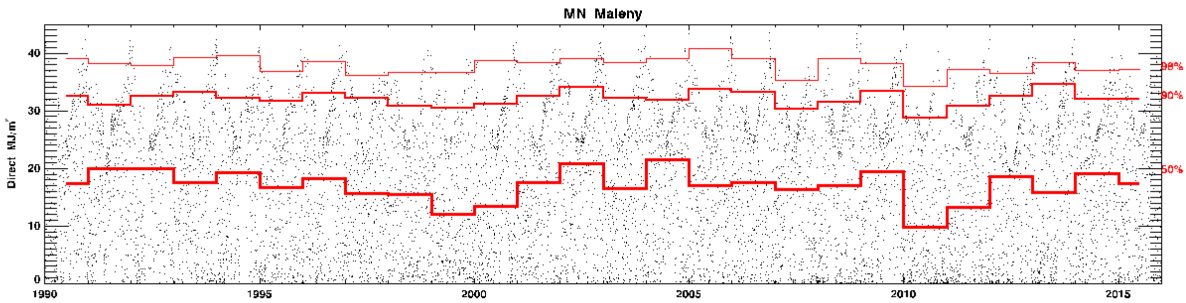
As above for **Maleny**.



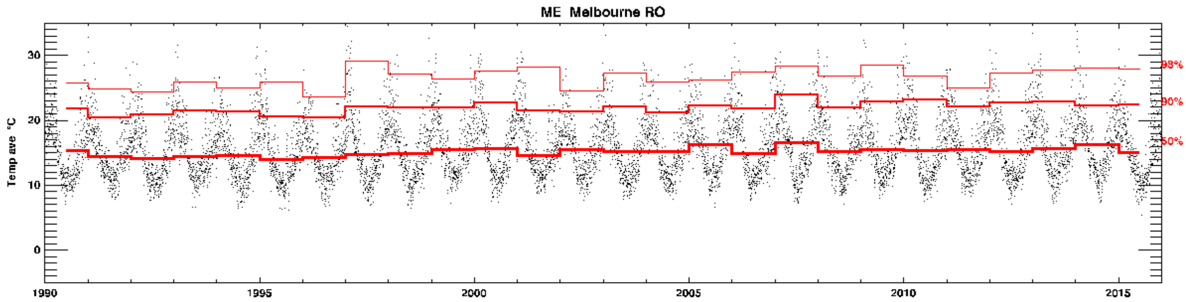


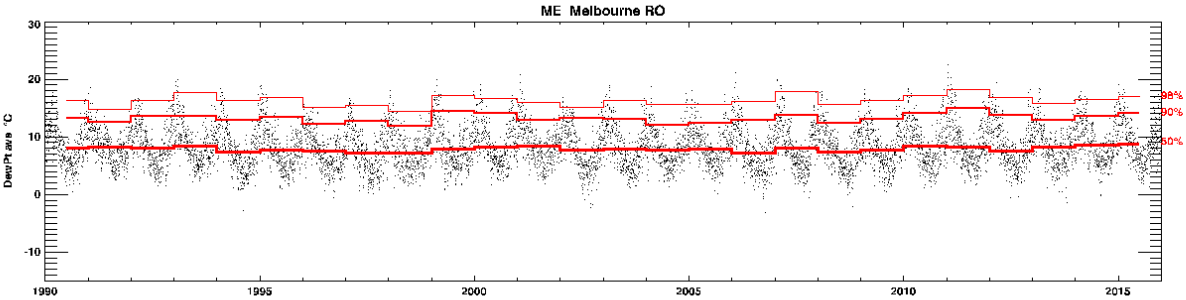


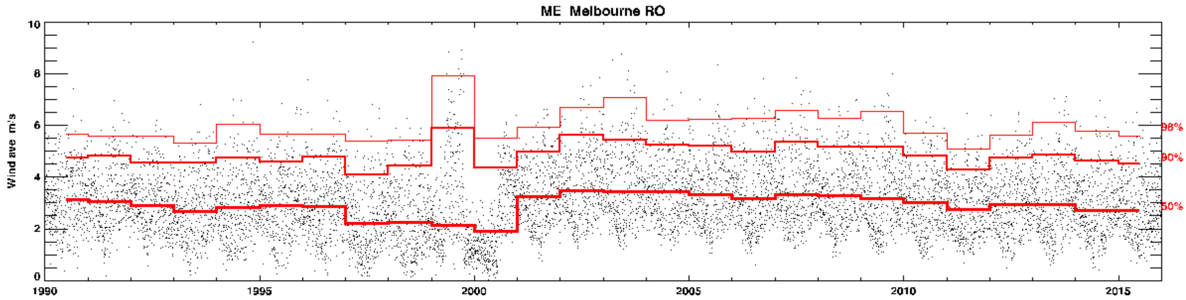


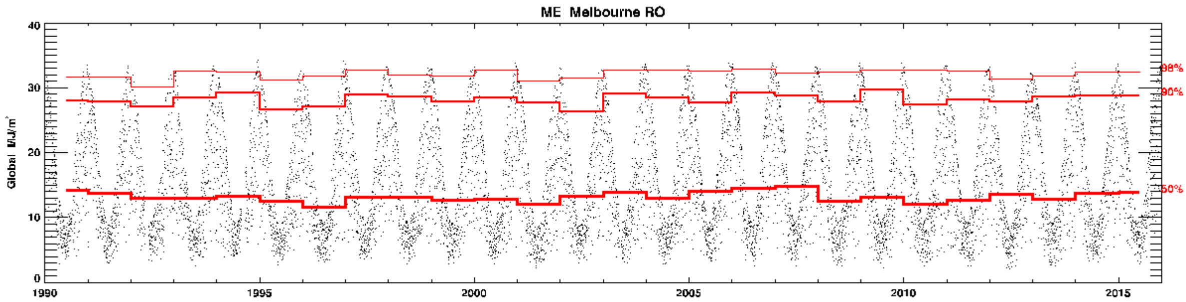


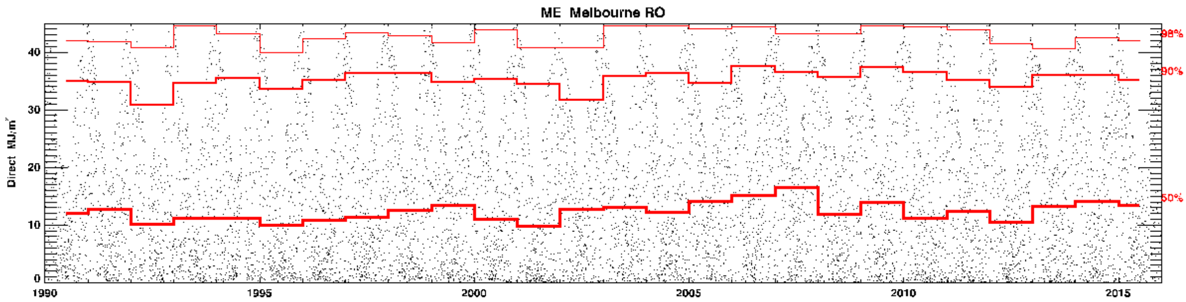
As above for **Melbourne RO**.



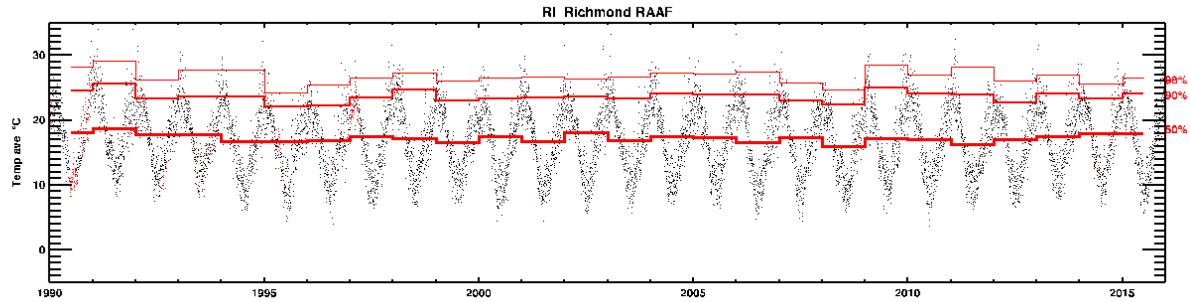


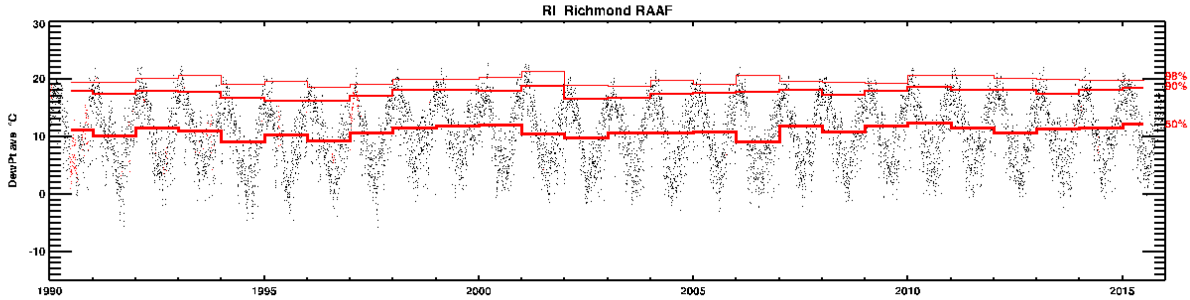


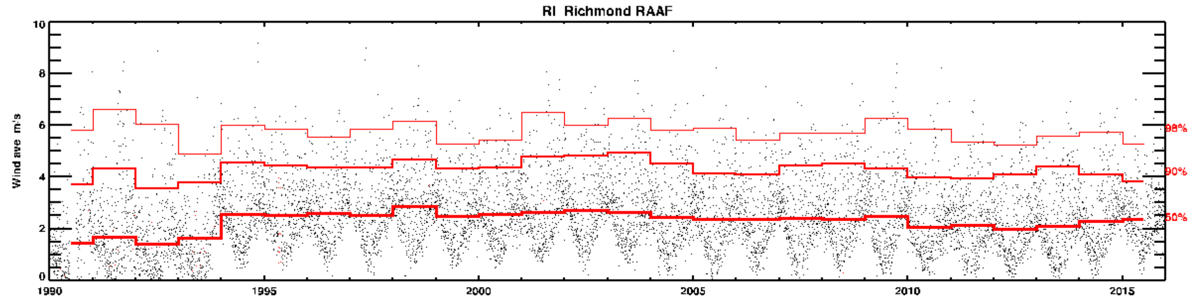


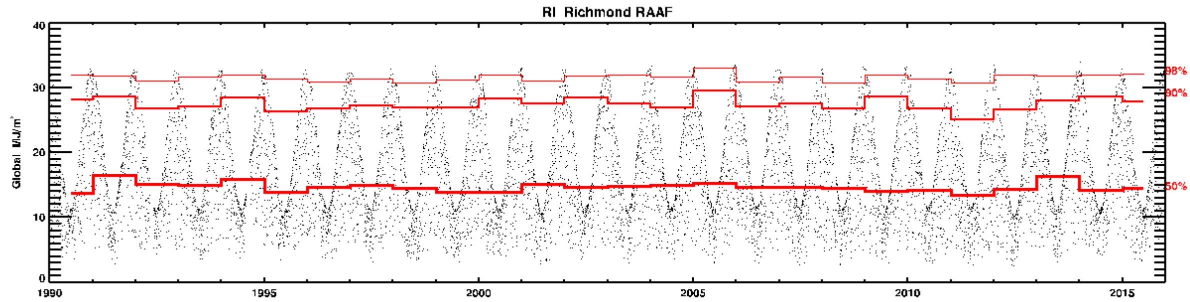


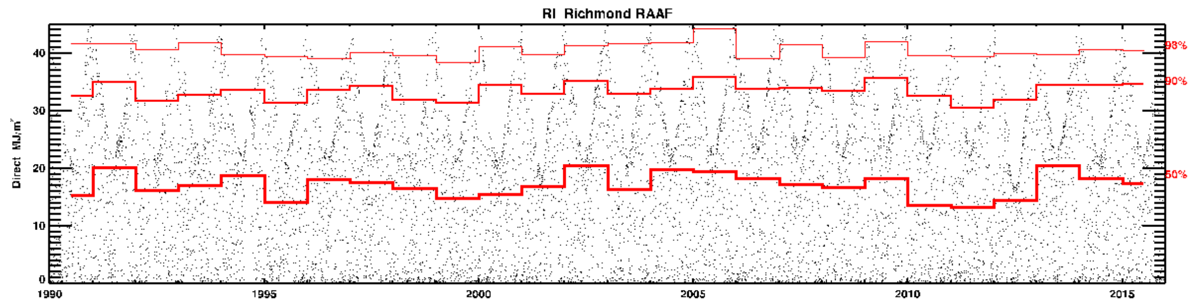
As above for **Richmond**.



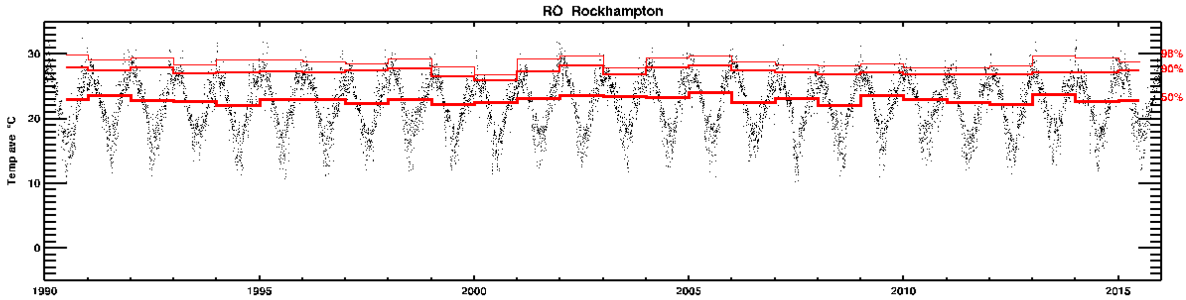


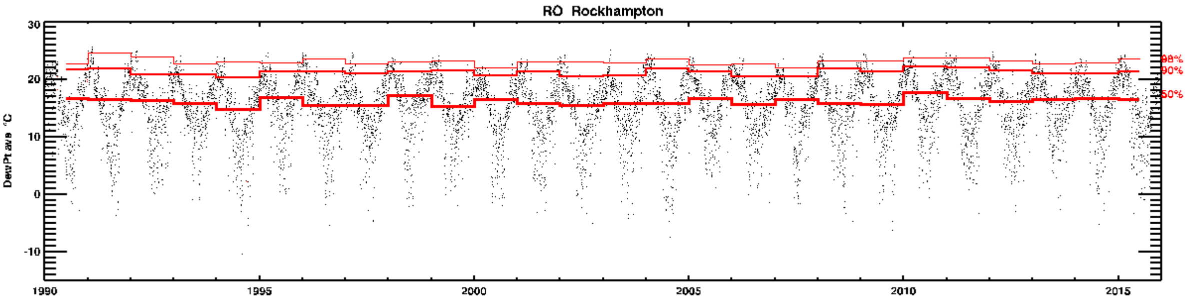


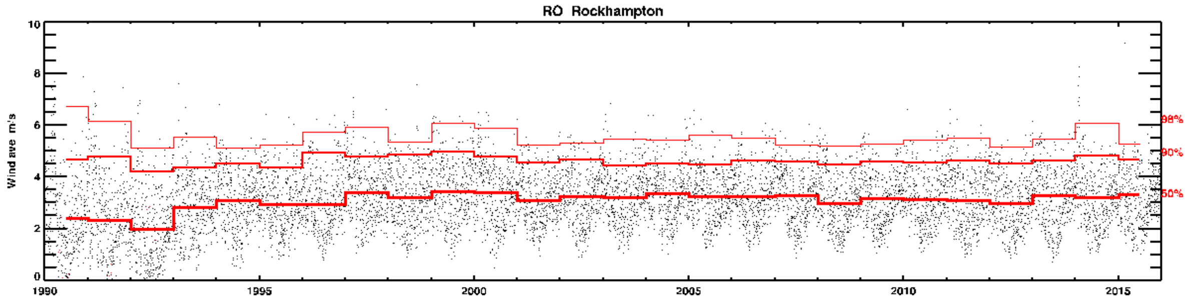


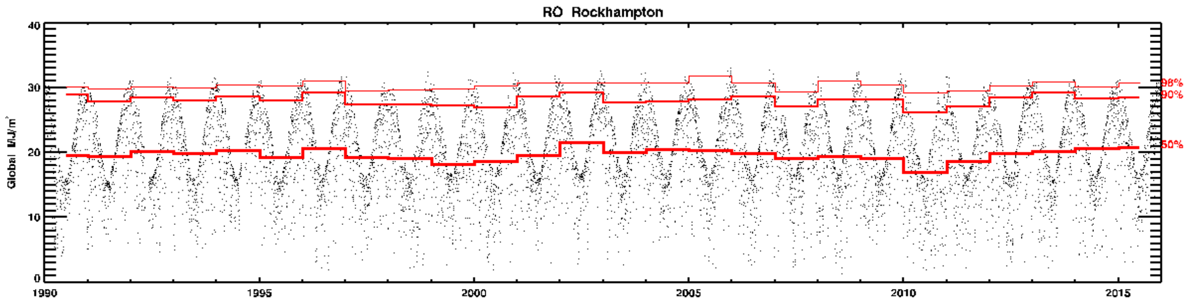


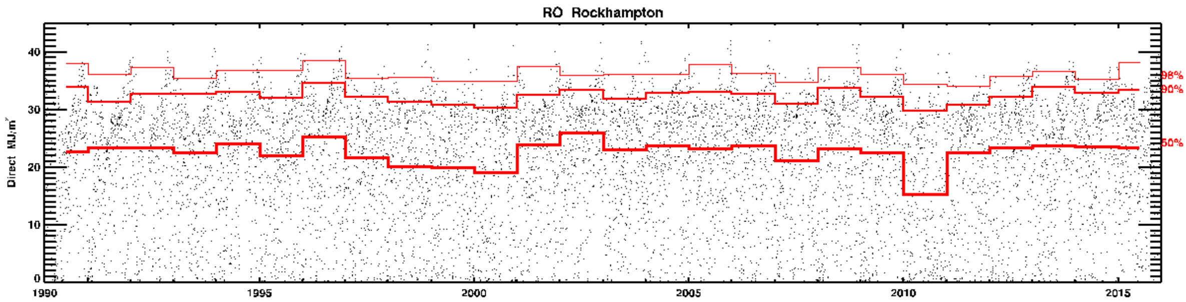
As above for **Rockhampton**.



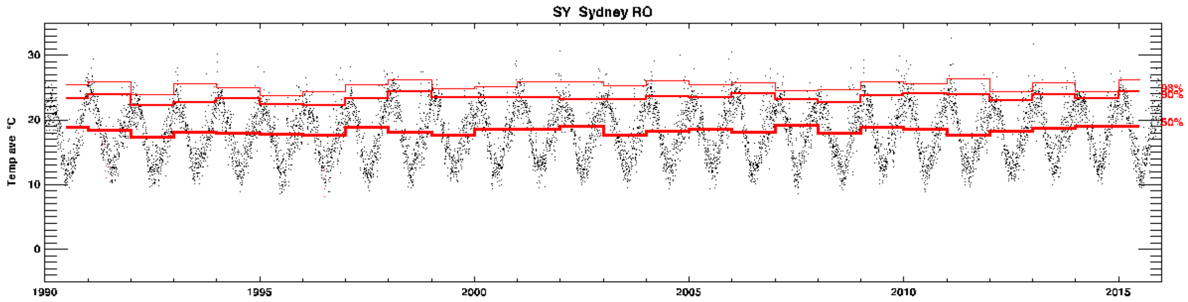


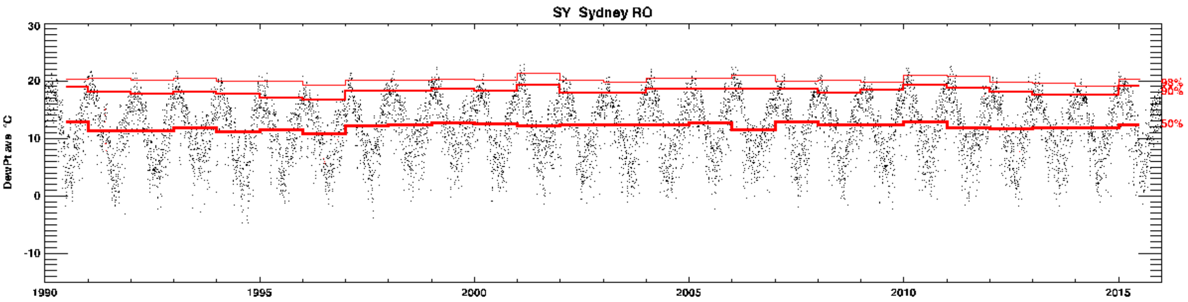


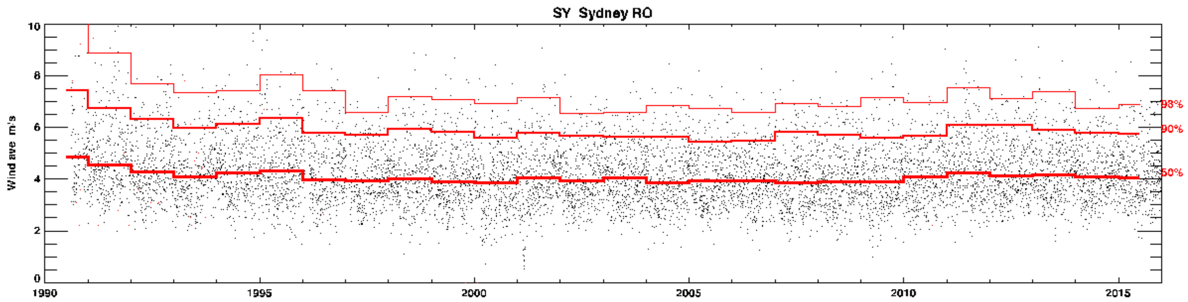


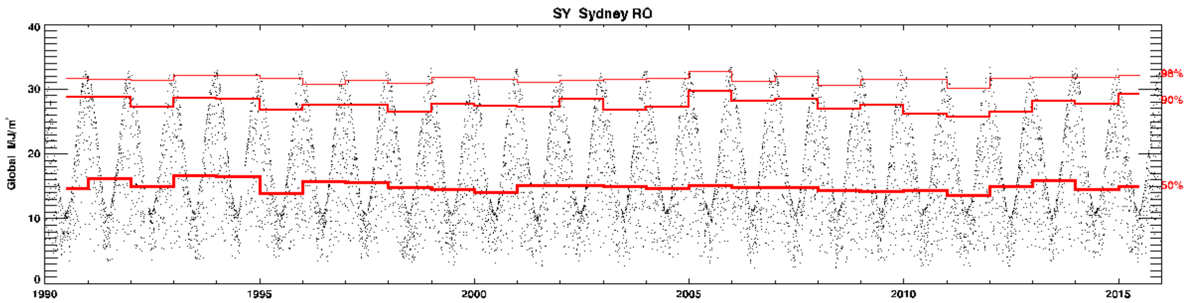


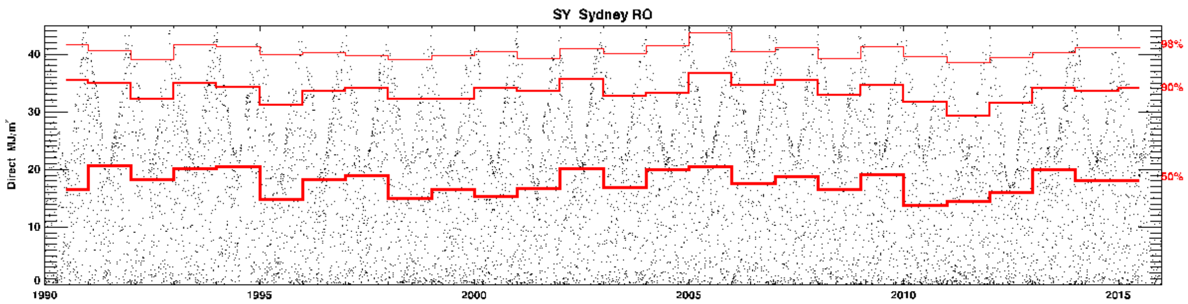
As above for **Sydney**.



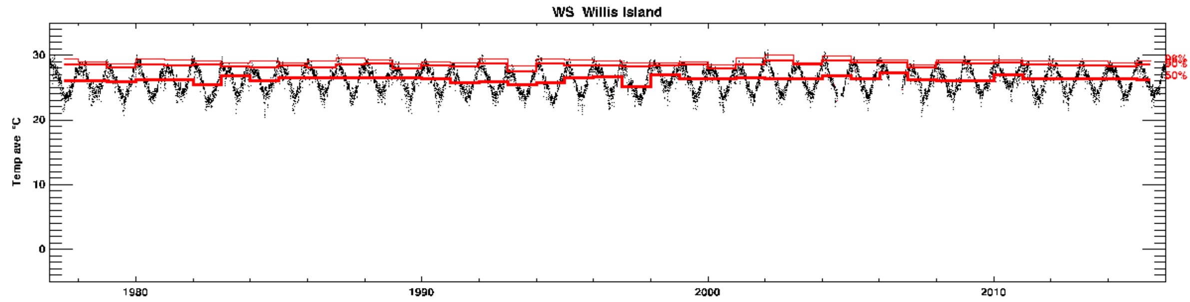


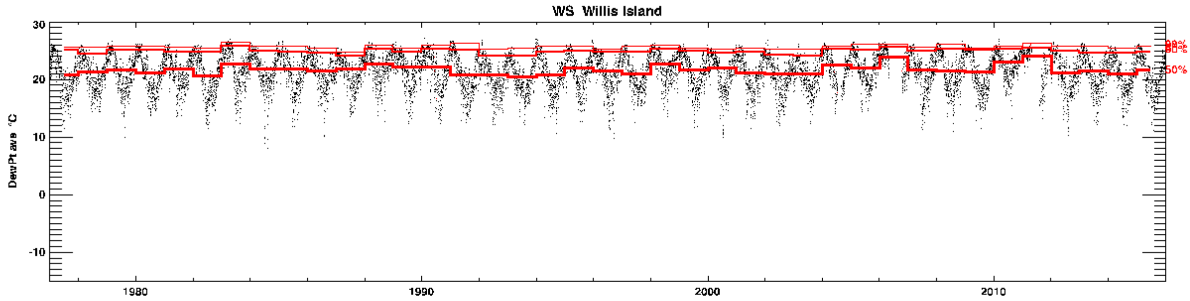


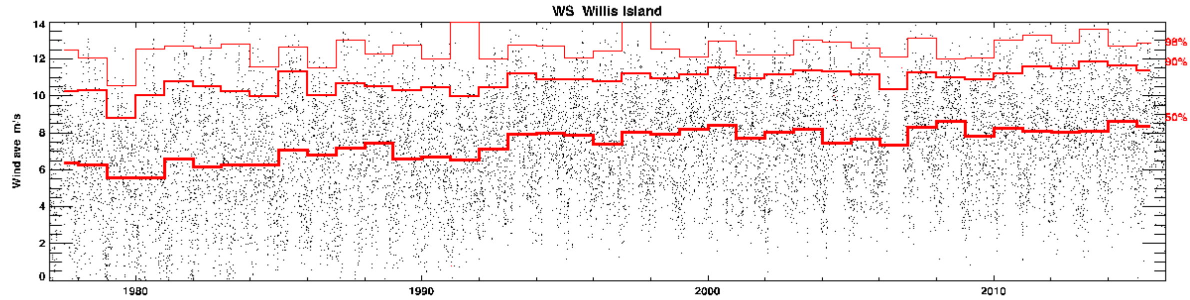


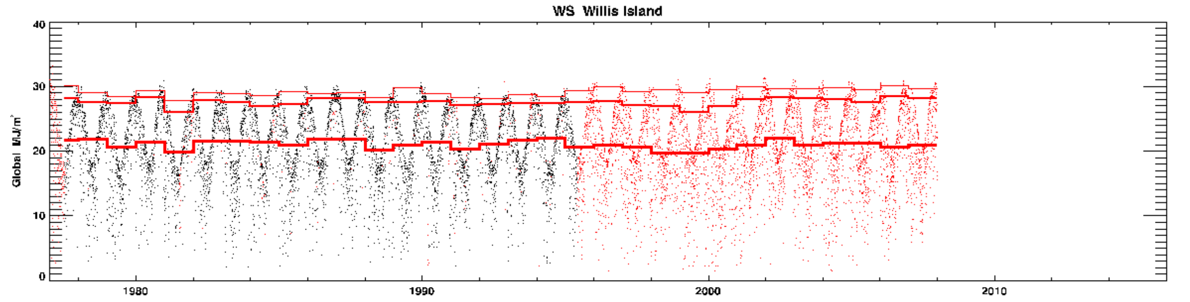


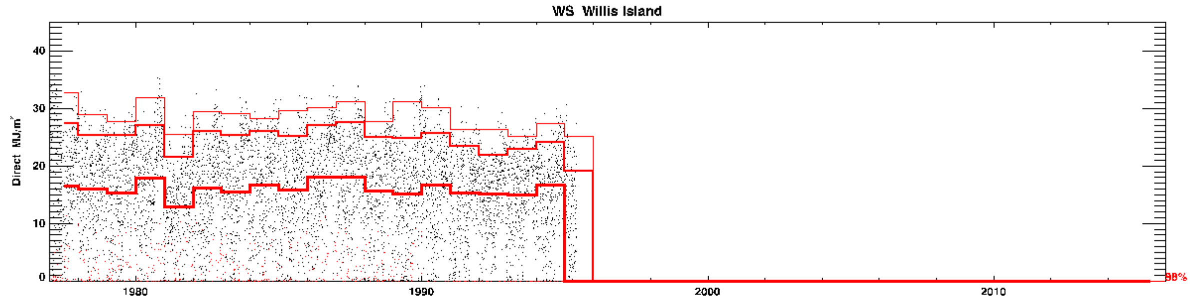
As above for **Willis Island**.











Appendix B Mathematical details of TMY2/RMY selection

Finkelstein-Schafer statistic

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

$$FS^{x}{}_{ym} = \frac{1}{n}\sum_{d=1}^{n}\left|D^{x}{}_{ym}(X_d) - D^{x}{}_{m}(X_d)\right|$$

where

X_d	is the value of parameter x on day d
$D^{x}{}_{ym}$	is the distribution of parameter x in month m of year y (black, Figure 21)
$D^{x}{}_{m}$	is the combined distribution of parameter x in month m (red, Figure 21)
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 RMY-A, -B, and -C weightings are listed in Table 4. Only the RMYA values are used in the present work.

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

$$FS_{ym} = \sum_{x} w_x FS^{x}{}_{ym}$$

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 five 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.

In past work, 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:

$$FSs^{x}{}_{ym} = \left|\frac{1}{n}\sum_{d=1}^{n}\left(D^{x}{}_{ym}(X_d) - D^{x}{}_{m}(X_d)\right)\right|$$

Referring to Figure 21, the true FS measures the mean absolute deviation of a month's distribution function from the combined distribution function, 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 value 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. Tests using possible measures other than FSs made only small changes to the order of preference among initially 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, and they provide a useful visual check of results. Each TMY2 or RMY comes from 120 plots like Figure 21 (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. That 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/RMY 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 (33rd and 67th 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 21. The combined distribution of run lengths enables evaluation of each month's distribution, as previously, with an FS-type statistic, FSr say.

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

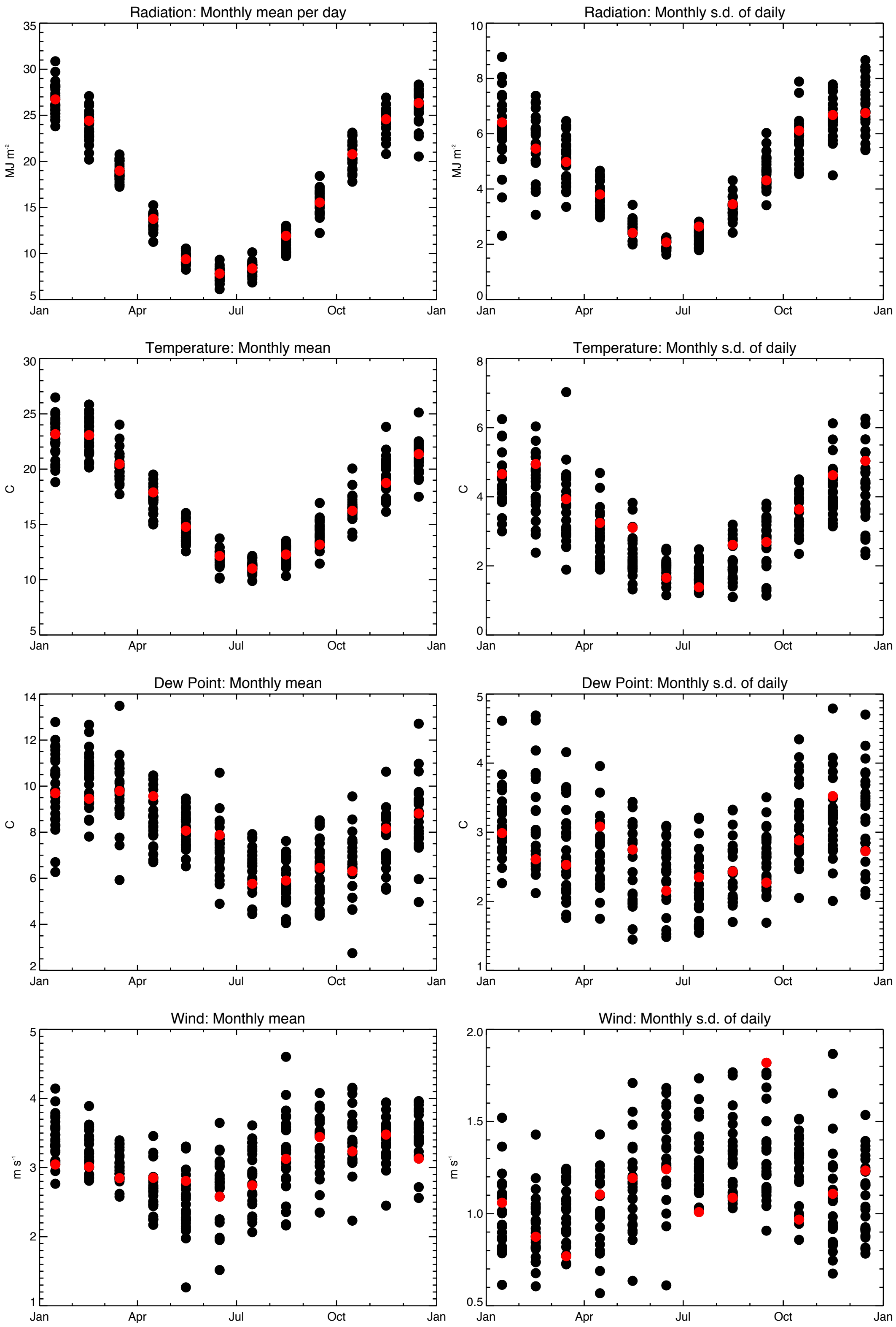
$$N_{ym}(l) = \sum_t w_t N^t{}_{ym}(l)$$

where

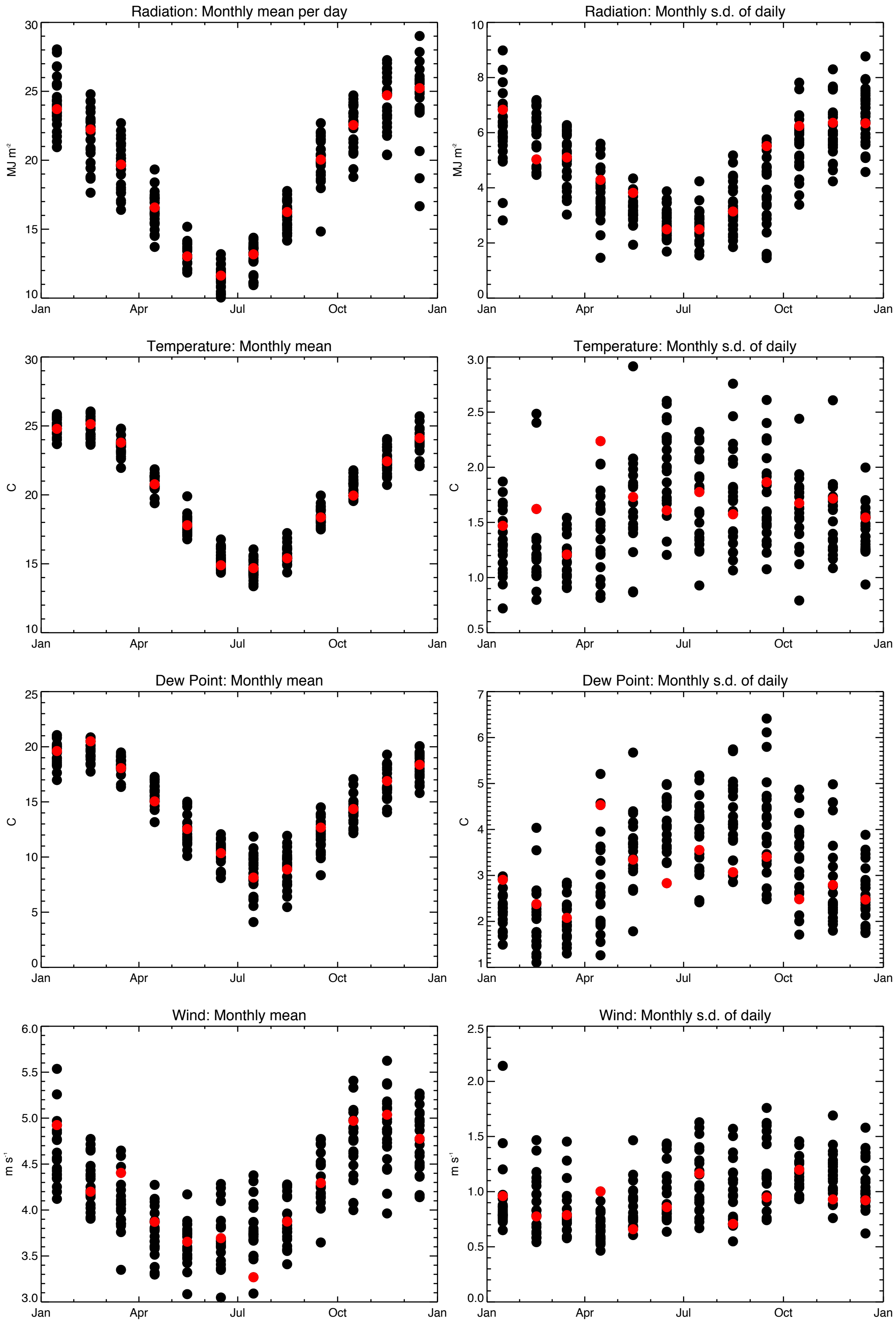
$N^t{}_{ym}(l)$	is the cumulative number of runs of length *l* in month *m* of year *y* for test *t* (parameter and tercile criterion)
$N_{ym}(l)$	is the weighted sum of the $N^t{}_{ym}(l)$
$\bar{N}_m(l)$	is the mean of $N_{ym}(l)$ across all years.

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 90th percentile for FSr.

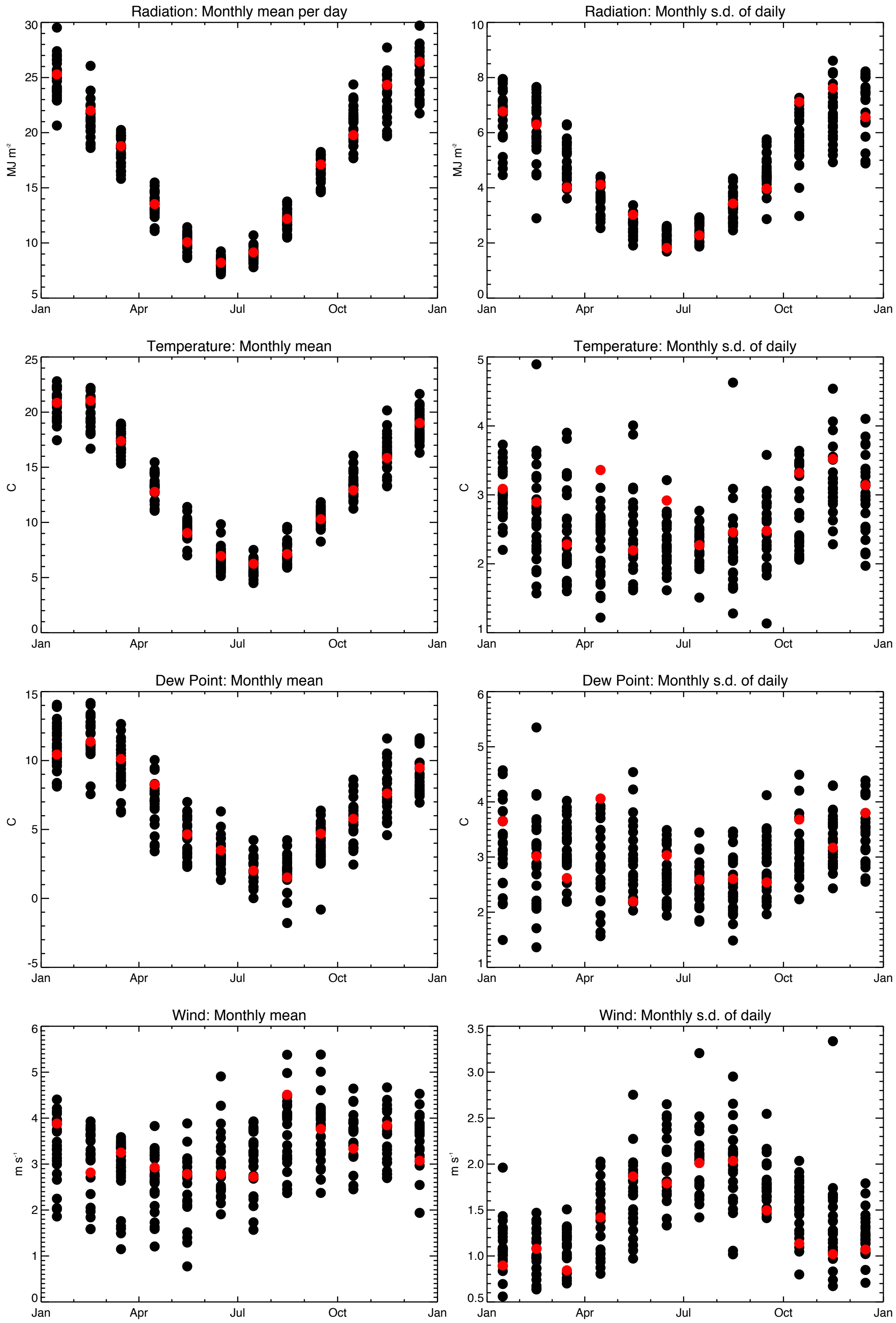
Appendix C Monthly statistics of selected sites and RMY



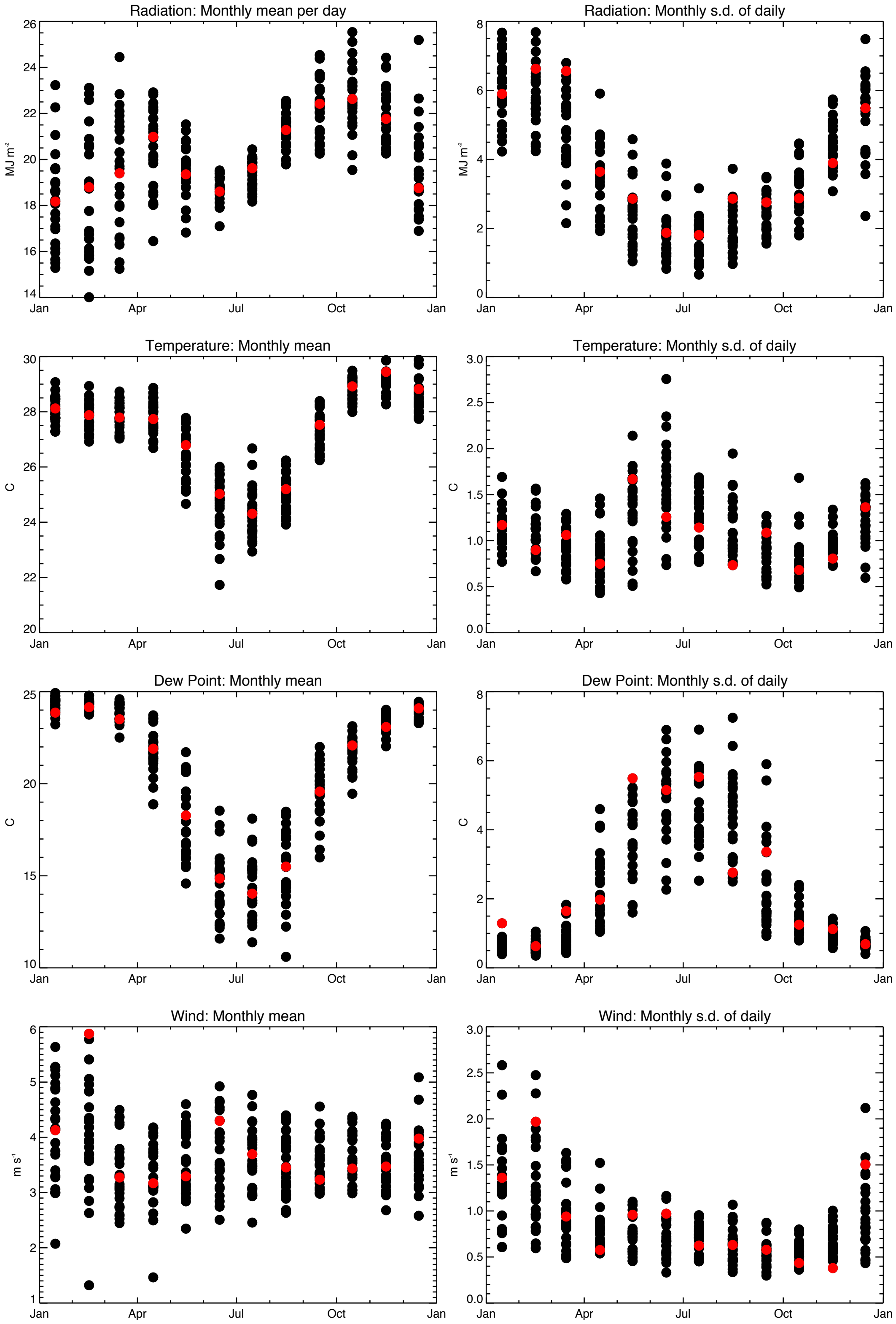
Monthly mean and standard deviation of daily climate parameters for **Adelaide**.
Red points indicate the selected months for the RMY.



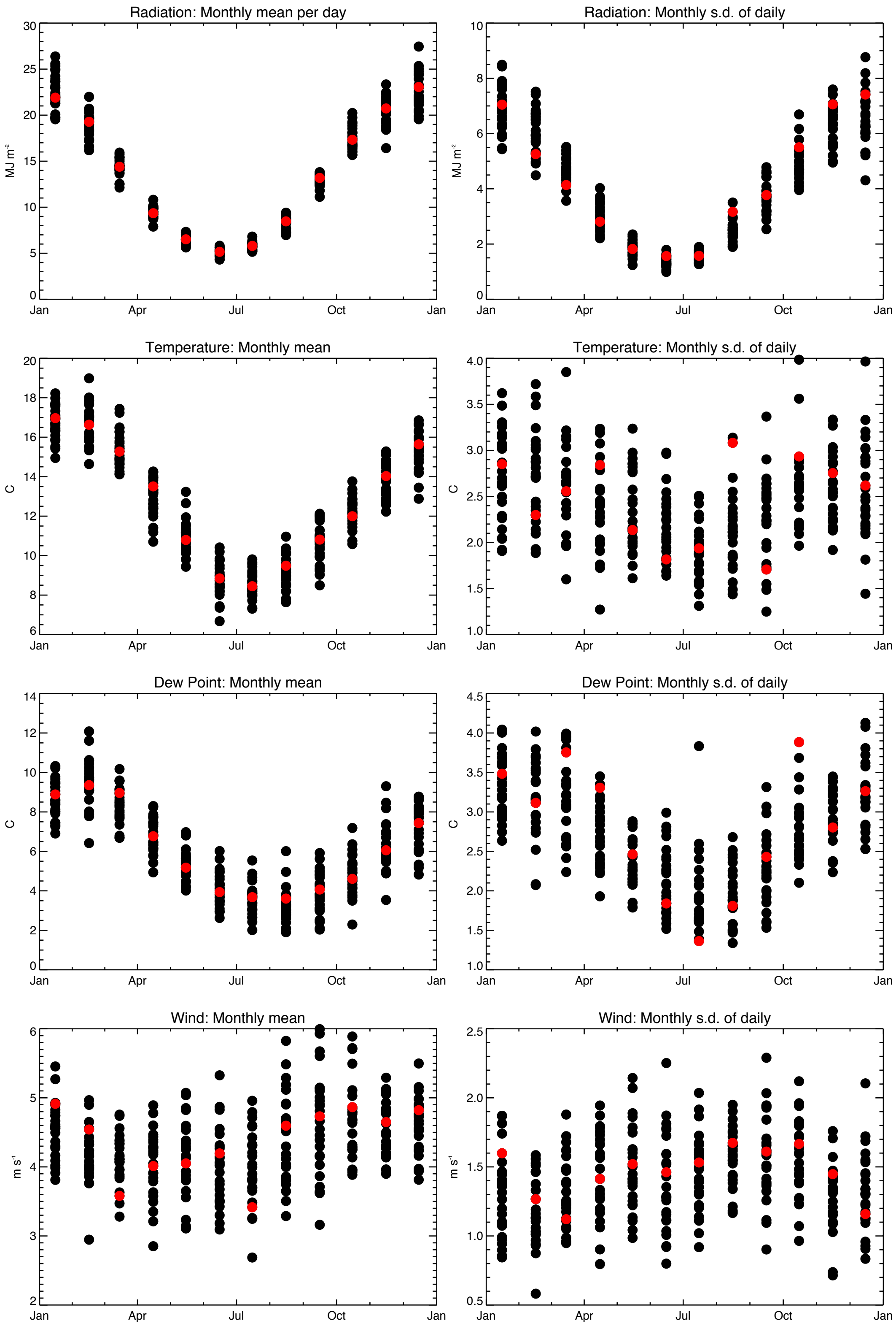
Monthly mean and standard deviation of daily climate parameters for **Brisbane**.
Red points indicate the selected months for the RMY.



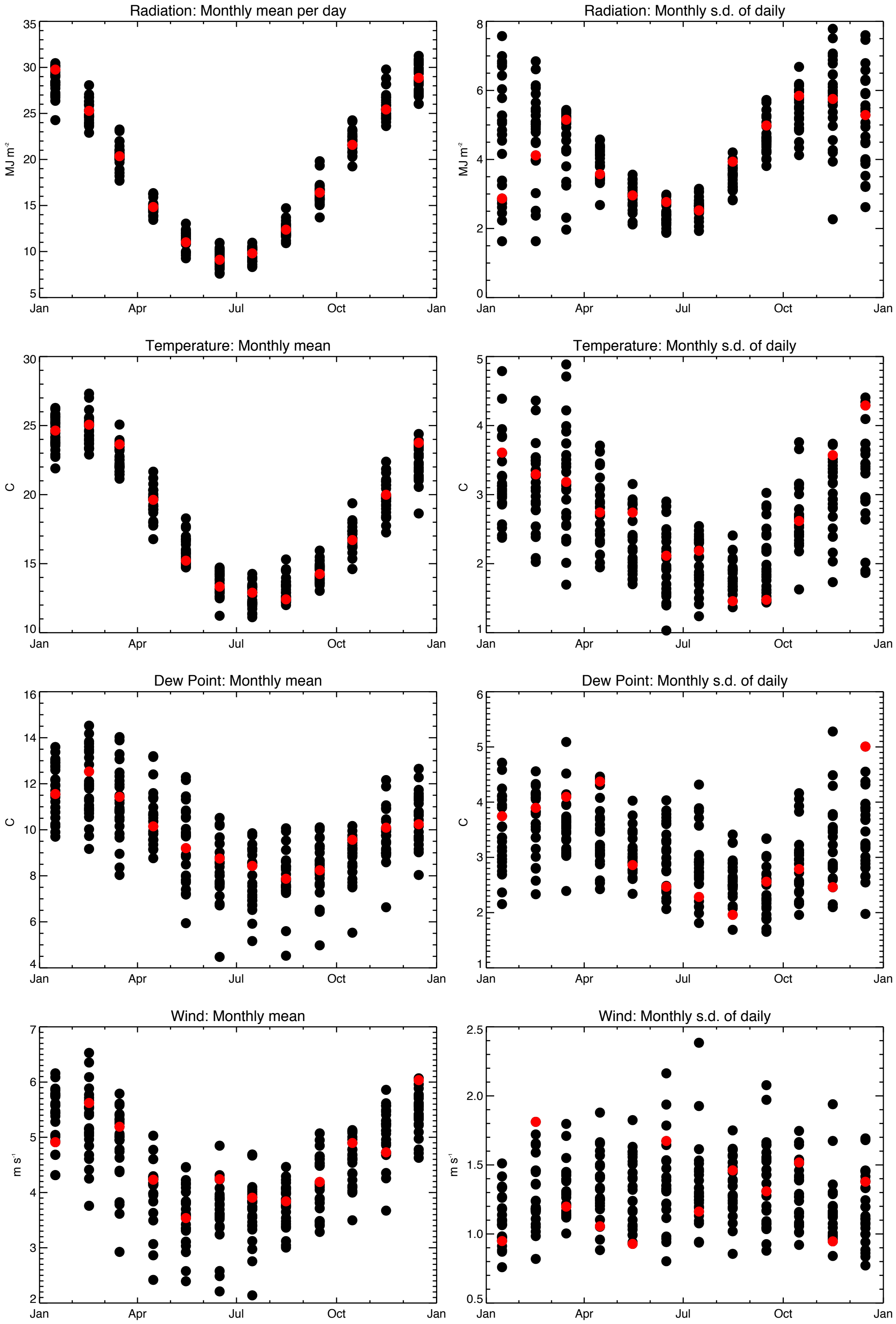
Monthly mean and standard deviation of daily climate parameters for **Canberra**.
Red points indicate the selected months for the RMY.



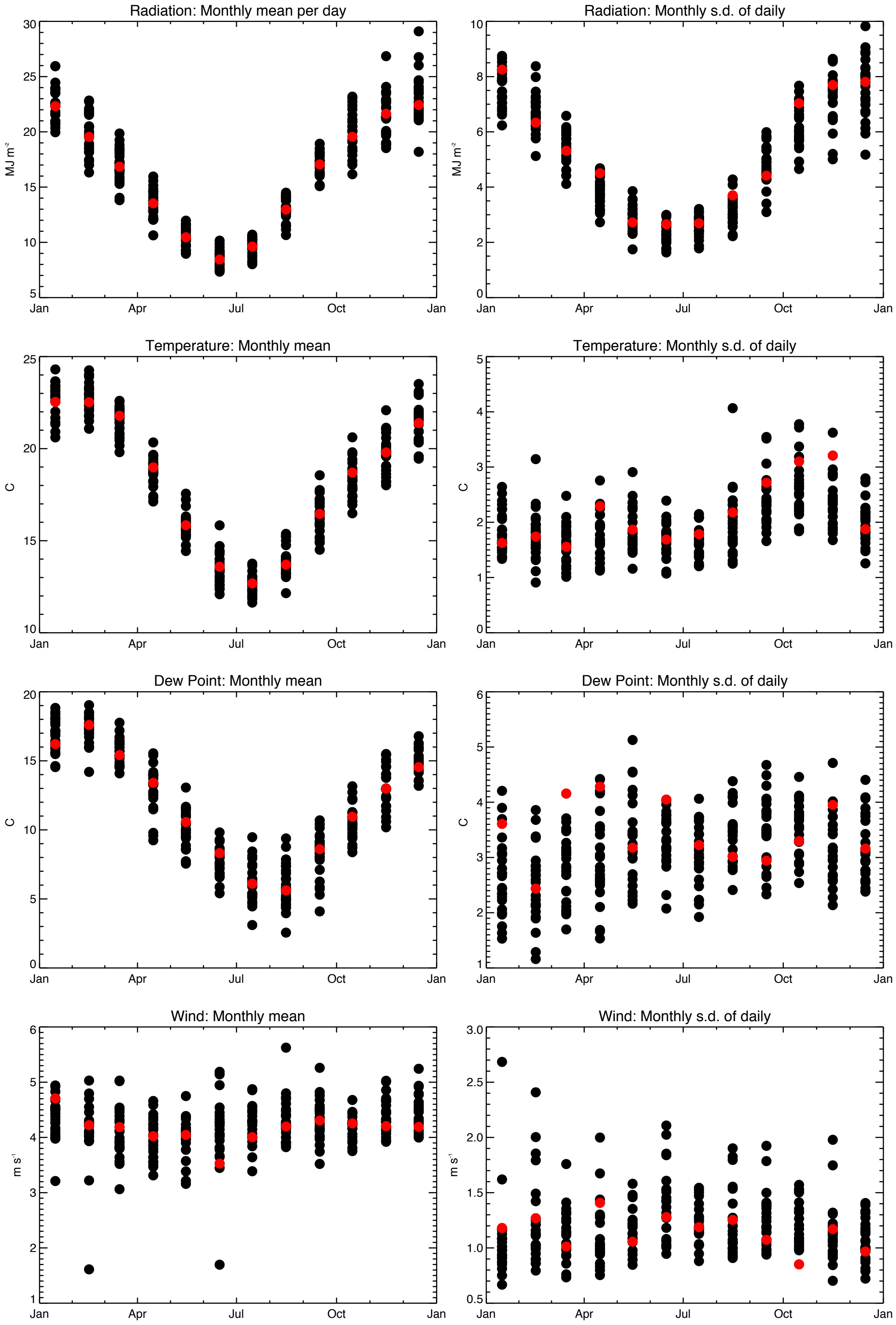
Monthly mean and standard deviation of daily climate parameters for **Darwin**.
Red points indicate the selected months for the RMY.



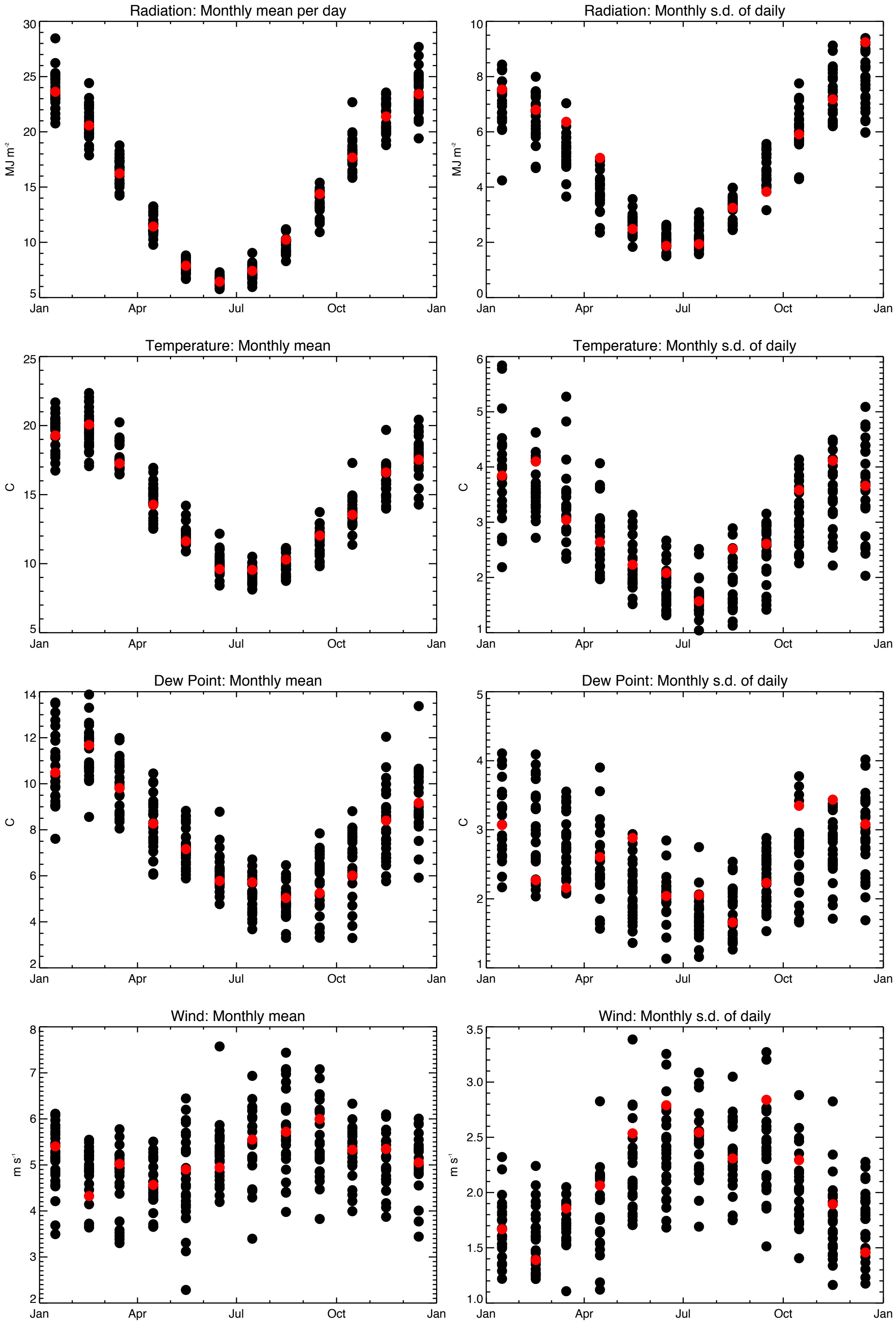
Monthly mean and standard deviation of daily climate parameters for **Hobart**.
Red points indicate the selected months for the RMY.



Monthly mean and standard deviation of daily climate parameters for **Perth**.
Red points indicate the selected months for the RMY.



Monthly mean and standard deviation of daily climate parameters for **Sydney**.
Red points indicate the selected months for the RMY.



Monthly mean and standard deviation of daily climate parameters for **Tullamarine**.
Red points indicate the selected months for the RMY.