How good are citizen weather stations? Addressing a biased opinion

Simon Bell, Dan Cornford and Lucy Bastin

Department of Computer Science, Aston University, Birmingham

Introduction

Currently over 1700 citizen weather stations (CWS) are observing and recording the weather across the UK. For comparison, the Met Office runs 250 or so land-surface stations in its professional Meteorological Monitoring System (MMS; Green, 2010). Tapping into this dense network of citizen observations could have benefits in many applications.

A CWS is defined as a weather station set up by a member of the public for whom the terms weather enthusiast, volunteer, hobbyist and amateur observer are fitting descriptions. Crucially, these stations are set up out of personal interest (or, in schools, for educational purposes) rather than because it is the owner's job. Increasingly the weather stations being used are consumer automatic weather stations, which are low cost and easy to install. Once connected to an internet enabled computer they can automatically submit observations to websites such the Met Office's weather observations website (WOW; http://wow.metoffice.gov.uk/), Weather Underground (http://www.wunderground.com) or the Citizen Weather Observer Program (CWOP; http://www.wxqa.com) for sharing. Most users submit their observations at subhourly intervals, most commonly every 5min. The majority of CWS are located in the owner's garden with most stations heavily clustered around urban and suburban areas.

The high temporal and spatial resolution of the freely available CWS' data potentially lends itself to many applications. The data could be fed into the data assimilation scheme of a high-resolution weather prediction model, used to post-process model output, analysed for urban heat-island studies (Steeneveld *et al.*, 2011; Wolters and Brandsma, 2012) or even used to estimate snow depths (Muller, 2013).

Using the CWS' data in the aforementioned applications raises concerns about data quality. Can CWS really compete with accurate equipment installed by professional organisations at well calibrated and exposed sites? Uncertainty about CWS' data can arise from any of the following five sources:

1. **Calibration issues** – a CWS sensor may not be perfectly calibrated. Perhaps it was biased before installation, or it has drifted over time.
2. **Design flaws** – often the design of a CWS makes it susceptible to inaccurate readings, particularly during certain weather conditions.
3. **Communication and software errors** – can produce gross errors as well as missing data.
4. **Metadata issues** – incomplete or inaccurate metadata make data interpretation difficult.
5. **Representativity error** – it is difficult to assess whether the CWS' observations represent a scale suitable for the application.

Here we present results from an intercomparison field study, where common models of CWS are collocated alongside professional equipment, allowing us to begin quantifying the magnitude of the bias and gross errors resulting from the first two sources, which we call instrumental errors. We also discuss how this field study has helped us to model the CWS bias. This is vital for quantifying such biases, and for structuring the framework of a quality control system that enables us to reduce biases and characterise residual uncertainty, allowing principled use of CWS data in scientific applications.

The field study

A year-long intercomparison field study was performed using seven CWS collocated alongside professional Met Office equipment at the University of Birmingham's 'Winterbourne No. 2' site (Figure 1). The site, located in Edgbaston (Birmingham), is part of the MMS network submitting minute-resolution data. The Met Office instruments installed at the site include a platinum resistance thermometer (PRT) and a Rotronic Hydroclip, both mounted within a passively ventilated Stevenson screen, a Munro R100 series 0.2mm tipping-bucket rain gauge and a Kipp and Zonen CMP11 pyranometer. The PRTs offer greater accuracy along with a more stable calibration in comparison with thermistors (Burt, 2012): all seven CWS we tested use thermistors.

Figure 1. The Met Office's Winterbourne No. 2 weather station. The site includes sensors operated by the Met Office, the University of Birmingham, and the seven CWS being tested as part of this study.

Weather – March 2015, Vol. 70, No. 3

RMetS
Royal Meteorological Society

Table 1

Summary of the seven CWS tested as part of this field study.

Station	*Station manufacturer*	*Station model*	*Price[a] (approximate)*	*Software used to download observations*	*Temporal resolution (min)*	*Time until memory full at this temporal resolution (days)*	*Rainfall increment (mm)*
VP2(1)	Davis Instruments	Vantage Pro2 FARS[b]	£890	WeatherLink	10	18	0.2
VP2(2)[c]	Davis Instruments	Vantage Pro2 FARS	£890	WeatherLink	10	18	0.2
Vue(1)	Davis Instruments	Vantage Vue	£390	WeatherLink	10	18	0.2
Vue(2)	Davis Instruments	Vantage Vue	£390	WeatherLink	10	18	0.2
WMR200	Oregon Scientific	WMR200	£350	Virtual Weather Station	10	291	1.016
WS2350	La Crosse	WS2350	£100	Heavy Weather	60	7	0.518
WH1080	Fine Offset[d]	WH1080	£70	EasyWeather	10	30	0.3

[a]Prices include accompanying software, but not mounting accessories such as tripods. Only the WMR200 comes with a mounting pole as standard. Prices include VAT.

[b]FARS stands for fan aspirated radiation shield.

[c]The VP2(2) had been in the field for approximately 1 year before installation at Winterbourne No. 2. All other stations were brand new.

[d]Fine Offset manufacture this station but it is frequently sold under many different brand names including Maplin, Watson, and Ambient Weather.

Henceforth these reference Met Office instruments will be referred to simply as 'MMS'. The MMS's instruments are calibrated on a regular managed cycle and abide by World Meteorological Organisation (WMO) standards (WMO, 2008); we assume therefore that they can be used as a well characterised reference against which the seven CWS can be verified. However, although this site will act as a good reference, the MMS's stations are not immune to problems. For example, passively ventilated Stevenson screens suffer from increased uncertainty at low wind speeds (Harrison, 2010), while tipping-bucket rain gauges, such as the Munro R100, can display biases when verified against standard manual rain gauges (Burt, 2012). Fortunately, the Munro R100 we used read just +1.1% higher than a Met Office MK II 'five-inch' manual rain gauge and +1.6% higher than another, newer, Munro tipping-bucket gauge, both collocated at the site. With this relatively small bias and a virtually complete annual dataset we can use its readings with reasonable confidence. The site is somewhat sheltered, and therefore unsuitable for Met Office wind measurements, but fortunately the University of Birmingham maintains a set of instruments at the site, including a 7m mast with an anemometer and wind vane manufactured by Vector Instruments. The site's sheltered nature is similar to that of many the CWS. The study took place from 1 September 2012 to 31 August 2013. Having envisaged that the type and magnitude of the CWS' bias would depend on synoptic conditions, which vary through the year, a full year's field study was undertaken.

The seven CWS comprised five different models of weather station, chosen because they are among the most popular automatic stations used by citizen observers (Bell *et al.*, 2013). Details of the stations are summarised in Table 1, with images of the sensor suites shown in Figure 2. The two stations of each of the Davis Instruments' Vantage Pro2(VP2) and Vantage Vue, and Oregon Scientific WMR200 were installed, with the aim of identifying biases and errors common to a particular model. The second WMR200 was decommissioned in early November 2012 when its wireless transmission began to interfere with that of the first station. We are confident that there was negligible interference before this point. Only a single Fine Offset WH1080 and La Crosse WS2350 were deployed because of fears of similar interference. With hindsight, the La Crosse instruments could have used wired communications and Jenkins (2014) used two Fine Offset devices simultaneously without issue. Like most CWS, every station comprised an outdoor sensor suite and an indoor electronic console to display and store the data. Observations were downloaded from the console to a laptop on a weekly basis. All CWS' and MMS's temperature and humidity sensors were mounted approximately 1.5m above grass. Although the rims of the MMS's rain gauges were roughly 30cm above grass, the heights of the CWS' gauges were set as recommended in their manuals, ranging between 1m for the WS2350 and WMR200 to 1.5–2m for the other CWS.

Results

Table 2 summarises some key statistics from the year-long field study. Below we discuss each weather variable in turn, namely: air temperature, humidity and dew-point temperature, and rainfall. Because pressure variations are well captured by the MMS network, and CWS' wind measurements will often be too localised for most applications, they are not examined in detail.

Temperature

When the air-temperature measurements from the seven CWS were verified against the MMS's measurements there were significant biases (Table 2), with clear diurnal and seasonal patterns (Figure 3). The pattern is dictated by the hours of daylight, with changes in the magnitude, and sometimes the sign, of the bias between day and night.

The Davis VP2 and Vue stations show the closest agreement with the MMS's PRT, all

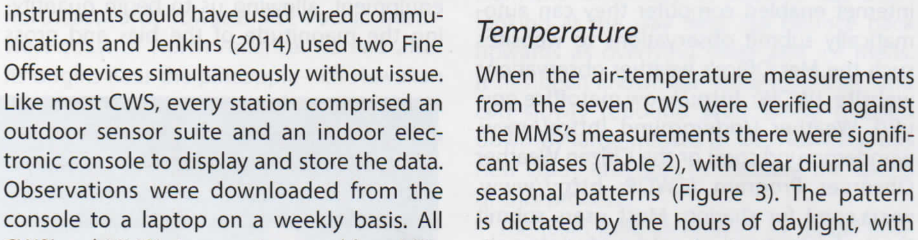

Figure 2. Citizen stations tested at the Winterbourne No.2 field site.

RMetS Royal Meteorological Society

with a relatively small mean bias and standard deviation. The two fan-aspirated VP2s show very similar results, both tending to read too warm, with overall average biases of +0.16 degC and +0.18 degC. Both show an increase in this warm bias at night and a decrease during the day. The two Vue stations, however, do not agree: Vue(1) tends to show a warm bias that is exacerbated in the afternoon, whereas Vue(2) consistently shows a cool bias, around −0.2 degC at night, which only changes to a warm bias in between late morning and late afternoon for the warmer months of the year, as shown in Figure 3.

The temperature bias of the other stations is more significant, each showing a warm bias that dramatically increases during the day, and leads to a positively skewed distribution of bias for these stations. The pattern of the warm bias shown in Figure 3 for the La Crosse WS2350 station is typical for all these three stations, with a warm bias that peaks just after midday. During summer months the warm bias is even more pronounced, occurring for more hours of the day owing to the extended hours of insolation. These warm biases are well over 1 degC and can climb over 4 degC for the WMR200 and WS2350. To put these biases into context, the summer average daytime urban heat-island measured in London rarely exceeds 1 degC (Wilby *et al.*, 2011). Without accurate bias correction for CWS it would be almost impossible to identify such an urban heat-island effect using some of the models. At night the performance of these three stations is much improved, for example the WMR200 and WH1080 both display a small mean bias (standard deviation) of 0.05(0.4) and 0.03(0.3) degC, respectively.

Figure 4 shows a temperature time-series plot for a typical day that highlights many of the patterns. For example, note the large daytime warm bias exhibited by the WS2350 station (and to lesser extents by the WMR200 and WH1080) and the warm bias of both Vue stations later in the day.

Humidity and dew point

All seven CWS tested exhibit significant relative-humidity biases when compared with the MMS's humidity sensor. It is important to note that there is some uncertainty associated with the MMS's humidity sensor, the Rotronics Hydroclip. Figure 5 shows a time series of the Vue(1) humidity observations minus the MMS's observations. The sudden step change in the bias range in mid-May is because the Rotronics Hydroclip was swapped for another as part of the site's calibration process. The Rotronics Hydroclip that ran over the first 8.5 months tended to read much wetter than the CWS' sensors during conditions of high humidity,

Table 2

Key statistics from the field study over the period 1 September 2012 through to 31 August 2013, except for relative humidity and dew point, where the statistics represent the period 16 May 2013 to 31 August 2013. The standard deviation of the difference is shown in parentheses next to the values of mean bias.

	Air temperature (°C)			Relative humidity (%)			Dew point (°C)	MSLP[a] (hPa)	Rainfall
Statistic / Station	Mean bias (day and night)	Mean bias (day time[b])	Mean bias (night time)	Mean bias (all conditions)	Mean bias (wet conditions, >90%)[c]	Mean bias (dry conditions, ≤90%)[c]	Mean bias	Mean bias	Absolute and percentage difference from the MMS yearly total of 842.4mm
VP2(1)	**+0.2** (0.2)	**+0.1** (0.2)	**+0.3** (0.2)	**+2.7** (2.9)	**−1.3** (1.2)	**+3.6** (2.3)	**+0.7** (0.6)	**+1.7** (0.3)	−83.4mm (−9.9%)
VP2(2)	**+0.2** (0.3)	**+0.1** (0.3)	**+0.3** (0.3)	**+0.4** (3.1)	**−2.2** (3.1)	**+1.0** (2.7)	**+0.3** (0.5)	**+1.2** (0.3)	+94.8mm (+11.3%)
Vue(1)	**+0.1** (0.3)	**+0.2** (0.3)	**+0.1** (0.2)	**+2.7** (2.1)	**−0.2** (0.9)	**+3.4** (1.7)	**+0.8** (0.7)	**+1.7** (0.6)	−22.6mm (−2.7%)
Vue(2)	**-0.1** (0.3)	**+0.0** (0.3)	**-0.2** (0.2)	**+3.9** (2.0)	**+1.1** (0.9)	**+4.5** (1.6)	**+0.8** (0.7)	**+2.9** (0.8)	−28.6mm (−3.4%)
WMR200	**+0.8** (1.3)	**+1.5** (1.4)	**+0.1** (0.4)	**−11.0** (6.3)	**−2.8** (4.1)	**−12.8** (5.2)	**−1.7** (1.4)	**+2.6** (1.5)	−43.8mm (−5.2%)
WS2350	**+0.9** (2.3)	**+2.1** (2.5)	**−0.5** (0.5)	**−1.4** (5.2)	**−1.3** (2.0)	**−1.4** (5.7)	**+0.9** (1.4)	**+1.9** (1.1)	−100.0mm (−11.9%)
WH1080	**+0.5** (0.9)	**+0.9** (1.0)	**+0.0** (0.3)	**+7.5** (3.2)	**+5.1** (1.9)	**+8.0** (3.1)	**+2.3** (1.3)	**+0.0** (0.7)	−203.4mm (−24.1%)

[a]Winterbourne No. 2 site does not have MMS's MSLP readings, instead observations from the Coleshill MMS site 16km away were used. The CWS' pressure readings were set to match the Coleshill reading at the start of the period, except for the WMR200 station for which the MSLP correction is based upon the elevation the user enters into the electronic console.

[b]Here the definition of daylight is when the MMS's global radiation sensor reads greater than $0Wm^{-2}$, therefore night time is when the reading is less than or equal to $0Wm^{-2}$.

[c]As measured by the MMS's humidity sensor.

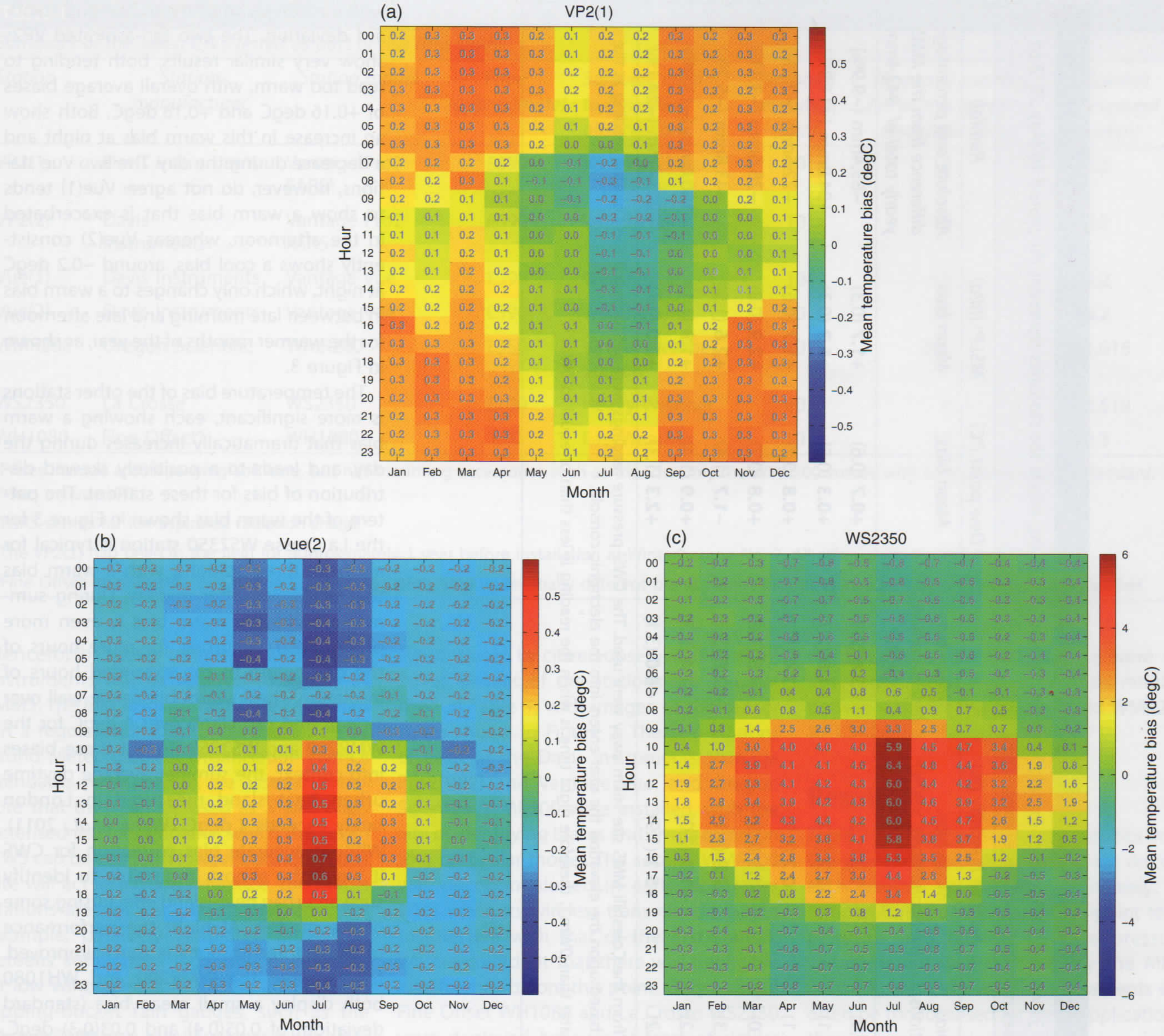

Figure 3. Mean temperature bias at different hours of the day (UTC) and months of the year for three of the CWS tested. (a) Davis VP2(1), (b) Davis Vue(2) and (c) La Crosse WS2350. Note the change in the colour scale for the final plot. The values written in grey are the mean bias of each cell.

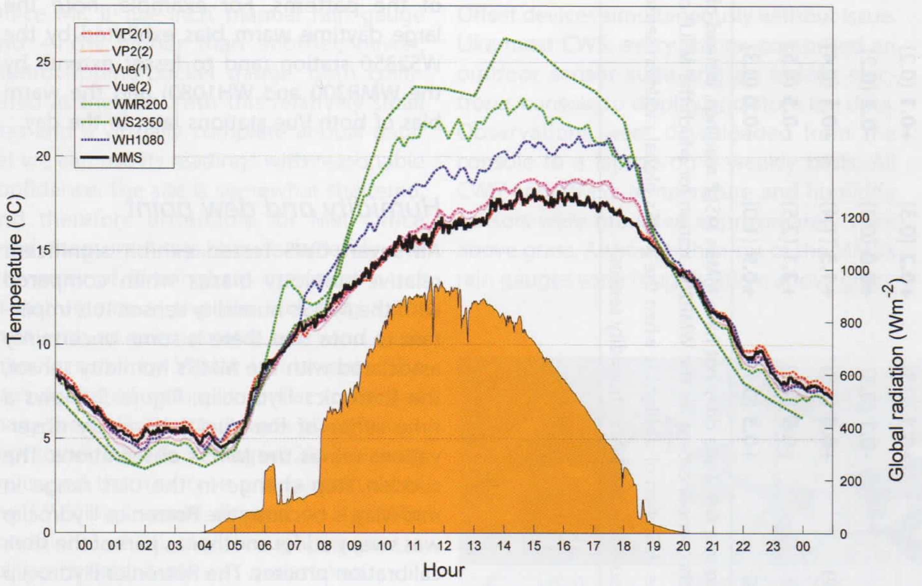

Figure 4. Time series plot of air temperature recorded by the seven CWS and the professional platinum resistance thermometer housed within a Stevenson screen for 26 May 2013. A time series of MMS global radiation is shown in orange.

remaining at 100% for several hours if not days (Figure 6). The CWS' observations would rarely read as high as 100%. This explains the large negative biases shown in Figure 5 over this first period. The second Rotronics Hydroclip showed no such tendency, exhibiting a much better agreement with a Vaisala humidity sensor run by the University of Birmingham at the site. In a separate field study, Ingleby *et al.* (2013) found that Rotronics Hydroclip sensors tend to drift by +1% to +2% per year at Met Office sites (although there is a lot of variability) and can be slow to recover from periods stuck at saturation. They estimate that an uncertainty of 2–3% for an operational Rotronics Hydroclip is achievable under best conditions. As the first Rotronics Hydroclip was deemed to have drifted wet, all following statistics and figures for relative humidity and dew-point temperature were produced using just the second Rotronics Hydroclip as the reference sensor.

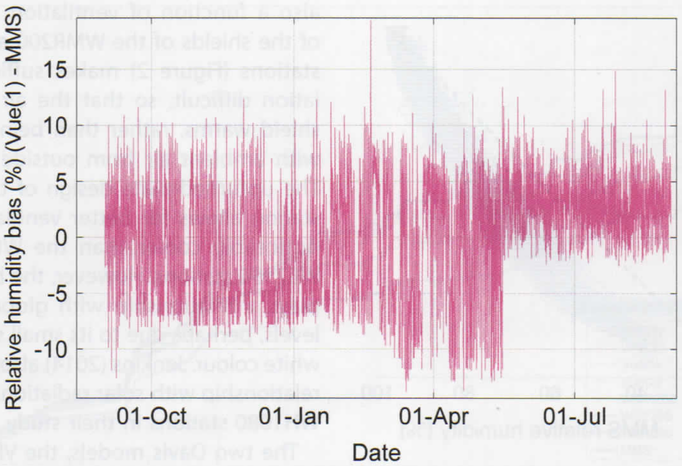

Figure 5. Time series of the Vue(1) station's relative-humidity bias, that is Vue(1) humidity - MMS humidity.

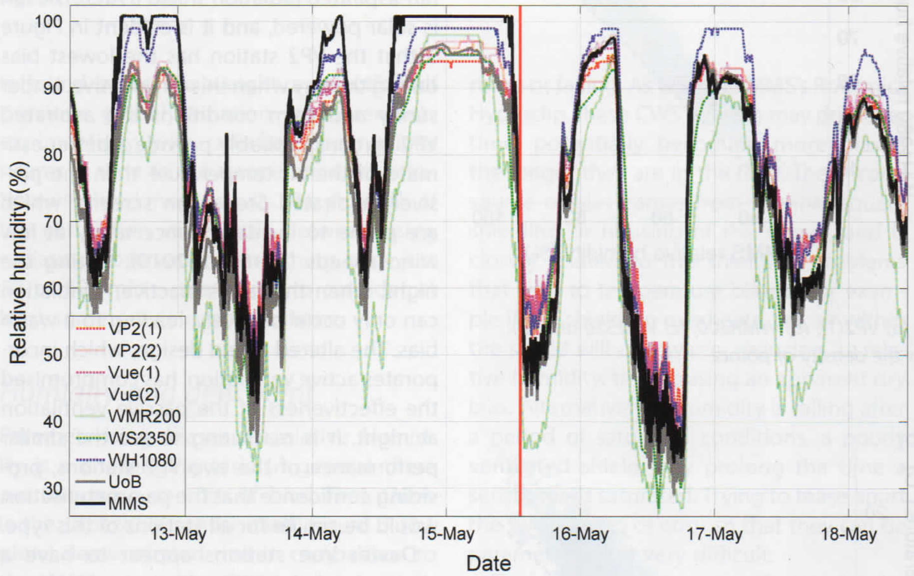

Figure 6. Relative humidity time series. Covers the period when the MMS's Rotronic Hydroclip was changed, as indicated by the red line. Note the addition of the University of Birmingham (UoB) Vaisala humidity observations.

Data from the humidity sensors of the seven CWS we tested have very different patterns to their bias. The Davis VP2(1), Vue(1) and Vue(2) stations all show a wet bias over the majority of the humidity range (Figure 7). For all three stations the mean bias is greater than 3% under drier conditions (less than 90%), but when the humidity is greater than 90% the VP2(1) and Vue(1) exhibit small dry biases. These findings agree well with those of Burt during his review of the VP2 (Burt, 2009) and Vue model (Burt, 2013). The VP2(2) behaves slightly differently: under drier conditions it does not overread to the same degree as the other Davis stations, but it underreads more during wet conditions, with a greater residual variance across the whole humidity range.

The WMR200 underreads across the entire humidity range, and dramatically so in drier situations, where it exhibits a mean bias of −12.8%. The WS2350 also tends towards a dry bias during drier situations, but with a less extreme mean of −1.4%. Between 70 and 90% (Figure 7) this switches to a wet bias. By contrast, the WH1080 has a large wet bias over the entire humidity range, with an overall bias of 7.5%.

As we anticipated some interaction of temperature and relative-humidity biases, we also considered dew-point temperature. The mean dew-point biases for all Davis stations were within 1 degC of the MMS. In agreement with other studies that found the VP2 station monthly means mostly within 1 degC of the reference sensor (Burt, 2009) and Vue station readings that were approximately 1 degC too high (Burt, 2013), both Vue stations in this study had a mean bias of +0.8 degC. The WS2350, with a mean bias of 0.9 degC was also within 1 degC of the MMS's sensor, but with a larger residual variance. The mean bias of the WMR200 and WH1080 stations was more significant, at −1.8 degC and 2.3 degC respectively (Figure 8).

The dew-point values of the CWS were derived by the station's electronic console, however, the difference due to their use of potentially different algorithms was virtually negligible, never exceeding 0.02 degC.

Rainfall

Figure 9 shows a plot of cumulative rainfall throughout the year-long study. All but the VP2(2) station measured totals less than the MMS's gauge. It is interesting that one VP2 station overread whereas the other underread, particularly as the VP2 model allows for calibration of the tipping buckets using a screw under each bucket. Before installing the VP2 stations they were calibrated in the laboratory so that on average both read within 2% of the truth. It is curious that, once in the field, they should deviate from the professional gauge by approximately 10%, and in different directions. When Burt (2009) tested a different VP2 station against a standard 'five-inch' gauge he found the annual total was just 1.8% higher, but the agreement was not consistent, with monthly differences ranging from −10% to +19%.

The Davis Vue stations show a very good agreement with the MMS's gauge and with each other, both undercatching by less than 4%. However, results by Burt (2013) show that this slight undercatch is not consistent throughout all Vue stations: as compared with the standard 'five-inch' gauge the annual total of their Vue stations was 9% too high. The WMR200 showed a reasonable agreement, undercatching by just 5.2% at the end of the period. The WS2350 and the WH1080 stations undercatch by greater amounts, with a yearly total of just 88% and 76% of the Met Office total respectively.

Parameterising station bias

Thus far we have seen that all seven CWS tested display some substantial biases. In this section we attempt to identify the factors determining these biases, with the aim of parameterising them.

Temperature

The most obvious pattern to the temperature bias is the difference between day and night, a result of changes in the radiative balance. For the WMR200, WS2350 and WH1080 stations the main driver of their daytime warm bias is the strength of incoming solar radiation. Figure 10 shows how these three CWS exhibit a greater warm bias with increasing levels of solar radiation. This relationship is largely linear, with some evidence of bias plateauing as global

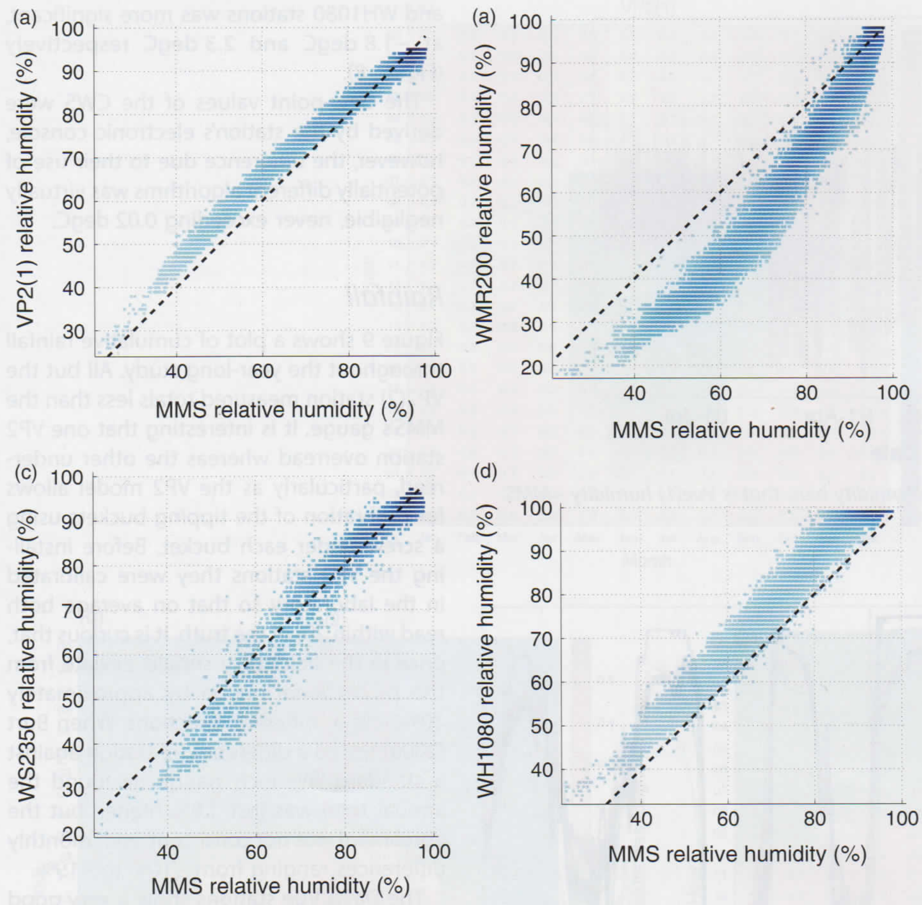

Figure 7. The CWS' versus MMS's relative humidity: (a) VP2(1), (b) WMR200, (c) WS2350 and (d) WH1080 stations. The darker the colour the greater the density of points.

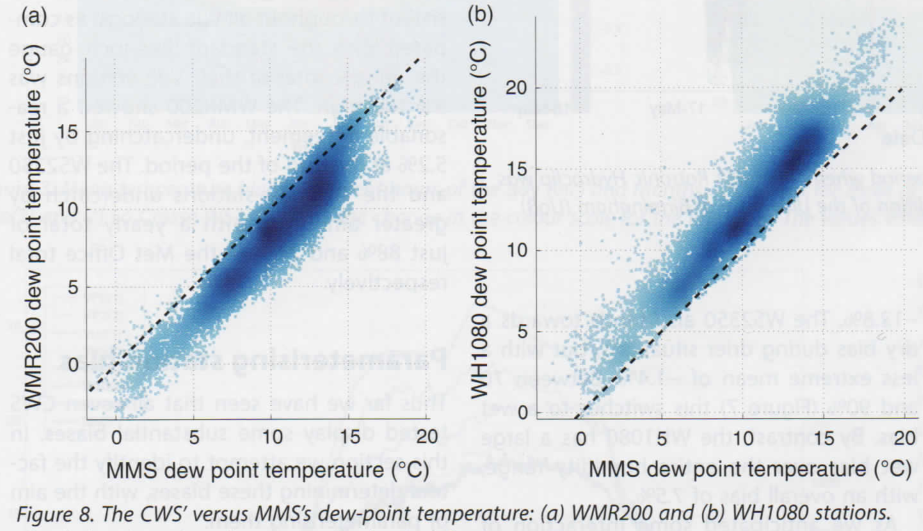

Figure 8. The CWS' versus MMS's dew-point temperature: (a) WMR200 and (b) WH1080 stations. The darker the colour the greater the density of points.

(direct + diffuse) radiation levels reach their highest levels.

Wind speed also appears to influence this relationship. Note that in Figure 11 the warm temperature bias of the WMR200 station during high solar radiation conditions is exacerbated when the wind speed is low.

All of the stations tested have some form of shielding to guard their thermistor from direct sunlight. Such shielding should also allow surrounding air to ventilate through, but it is apparent that some shields are more effective than others. Thermal imaging (Figure 12) shows that the WMR200 and WS2350 stations, which exhibit the largest biases under increased global radiation, exhibit the warmest colours on their shields. This illustrates that their thermistor shielding is prone to overheating under sunny conditions, which heats the air inside the thermistor housing, thereby increasing the sensed temperature. This overheating is also a function of ventilation: the design of the shields of the WMR200 and WS2350 stations (Figure 2) makes sufficient ventilation difficult, so that the air within the shield warms, rather than being refreshed with ambient air from outside the shield. The upturned-plate design of the WH1080 station allows for better ventilation and is noticeably cooler than the WMR200 and WS2350 stations. However, the bias still displays a relationship with global radiation levels, perhaps due to its small size and off-white colour. Jenkins (2014) also identified a relationship with solar radiation for the two WH1080 stations in their study.

The two Davis models, the VP2 and Vue, show the coolest colours in the thermal images, and their relationship with radiation is somewhat different. The VP2 stations were the only model we tested that included a fan-aspirated radiation shield (FARS).The fan is solar powered, and it is evident in Figure 3 that the VP2 station has the lowest bias during the day when this fan is active. Under sunny and calm conditions the aspirated VP2 stations probably provide a better estimate of the air temperature than the passively aspirated Stevenson screens, which are prone to increased uncertainty at low wind speeds (Harrison, 2010). During the night, when the fan is inactive, ventilation can only occur passively, leading to a warm bias. The altered shield design, which incorporates active ventilation, has compromised the effectiveness of the passive ventilation at night. It is reassuring to see the similar performance of the two VP2 stations, providing confidence that the parameterisation would be similar for all stations of this type.

Davis's Vue stations appear to have a stronger relationship with outgoing longwave radiation than shortwave radiation. As the Vue station's radiation shield is mounted underneath its main body it is well shielded from incoming solar radiation, although its effectiveness may be compromised when the solar angle is low. It is when the land surface has warmed and outgoing radiation peaks, around mid-afternoon, that the station shows the greatest warm bias (Figure 3). Unfortunately, longwave radiation is not commonly measured at MMS's stations, so a proxy variable may have to be used to parameterise the Vue station's temperature bias. The bias in the Vue station's temperatures demonstrated a stronger correlation with grass temperature than with global radiation.

So far we have only looked at the correlation between simultaneous CWS' temperature bias and global radiation measurements. This assumes that the impact of changes in radiation on the temperature bias is instantaneous. When a lagged impact is considered this correlation often improves. For example, for the WMR200, WS2350 and WH1080 stations, a better correlation

RMetS
Royal Meteorological Society

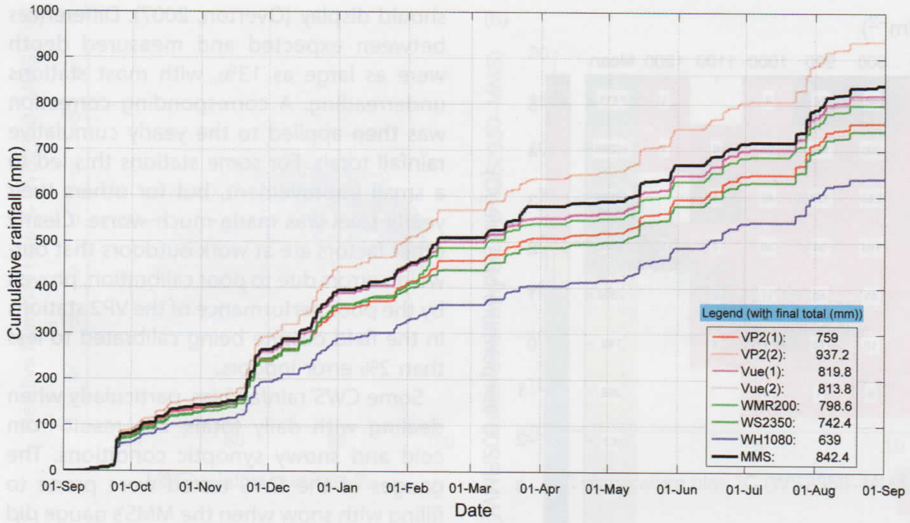

Figure 9. Cumulative rainfall totals of the seven CWS throughout the year-long study. Data from the professional Met Office gauge are shown by the black line. The final totals are displayed in the legend.

is found when we weight a selection of previous global radiation measurements exponentially, with radiation observations nearer in time to the temperature observation weighted more heavily. The strength of the correlation varies little between using radiation observations stretching back just 30 or 120 min – the key is to consider at least some previous observations.

Humidity and dew point

Parameterising biases in relative humidity is a challenging task. In general there are two main sources of bias. First, there is the capacitive sensor itself. Figure 7 (1:1 plots) demonstrated that in comparison to the MMS's sensor the CWS have potentially large calibration errors. The magnitude and even the sign of the bias will often change depending on the humidity. Bias may also be induced from hysteresis, when the sensor's response to a change in humidity varies depending on whether the humidity is rising or falling. As with the MMS's Rotronics Hydroclip, these CWS' sensors may drift over time, potentially becoming more biased the longer they are in the field. The second source of bias comes from the inadequate shielding or housing of the sensor, and is closely related to the shielding problems that lead to temperature biases. For example if the shielding overheats, the air within the shield will also warm, reducing its relative humidity, thus causing an apparent dry bias. Alternatively, if humidity is falling after a period of saturated conditions, a poorly ventilated shield may prolong the time a sensor reads saturated. Trying to tease apart the two sources of error so that they can be parameterised is very difficult.

Figure 13 shows a series of plots for the WS2350 station that can help us decipher the source of its humidity and dew point bias. It is apparent in Figure 13(a) that the relationship between the humidity bias and the MMS's humidity is somewhat unclear. Figure 13(b) plots the humidity bias against temperature bias. The majority of points display a slight cold and wet bias, however, there is a long tail where warm temperature biases are associated with a dry humidity bias. This may be caused by increased temperatures within the sensor housing, relative to the surrounding air. Bias in the dew-point temperature will be inherited from both the humidity bias and the temperature bias. Unsurprisingly Figure 13(c) shows that a warm temperature bias will lead to a subsequent warm dew-point bias, however, the relationship is not perfectly linear. As expected the dew point is too high when the humidity is wet biased. However, it is also too high when humidity is dry biased, perhaps because a warm temperature bias has not only caused a subsequent high dew-point bias but also a dry humidity bias when the sensor housing overheated. It is worth noting that we have chosen to plot just one station here: the plots for the other stations can look very different, further indicating that the parameterisations must be learnt for each of the CWS separately.

Rainfall

Six of the seven CWS tested displayed yearly rainfall totals lower than the MMS's rain gauge. One explanation for this could have been that, at 1–2m above the ground, the CWS' rain gauges are mounted higher than the MMS's gauge at 30cm, as such they experience higher wind speeds, which can lead to undercatch (Guo *et al*., 2001). In this study we looked for a relationship between wind speed and the daily CWS' rainfall totals as a proportion of the daily MMS's total. However, no obvious relationship was found, either for average wind speed taken at times of measured rainfall, or for daily average wind speed. Measurements from both the CWS's anemometer and the Vector Instruments anemometer, mounted at 7m, were used separately. Possible

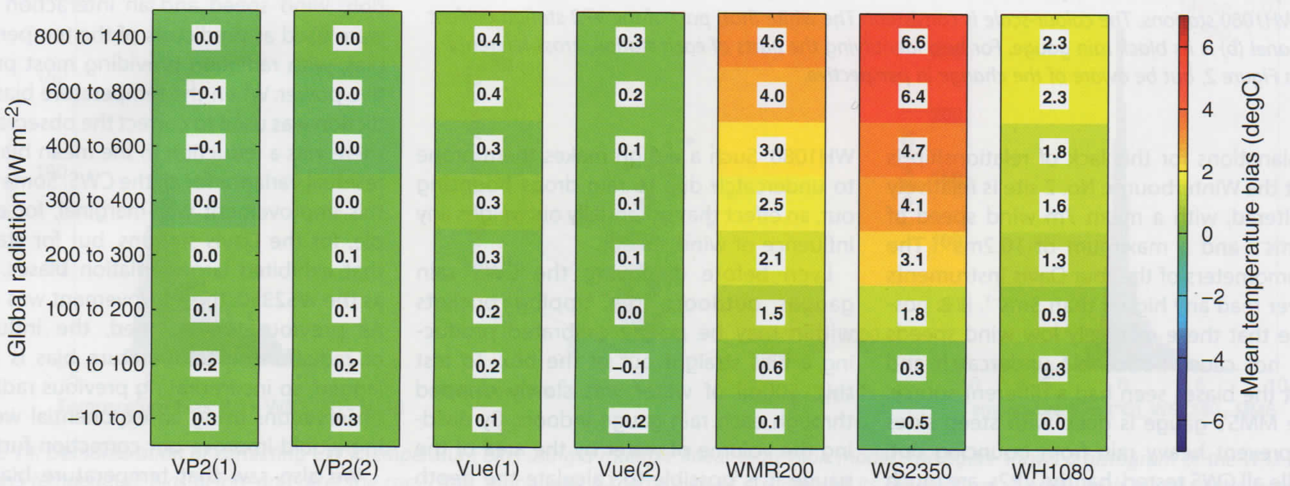

Figure 10. Temperature bias as a function of global radiation levels for the seven CWS tested.

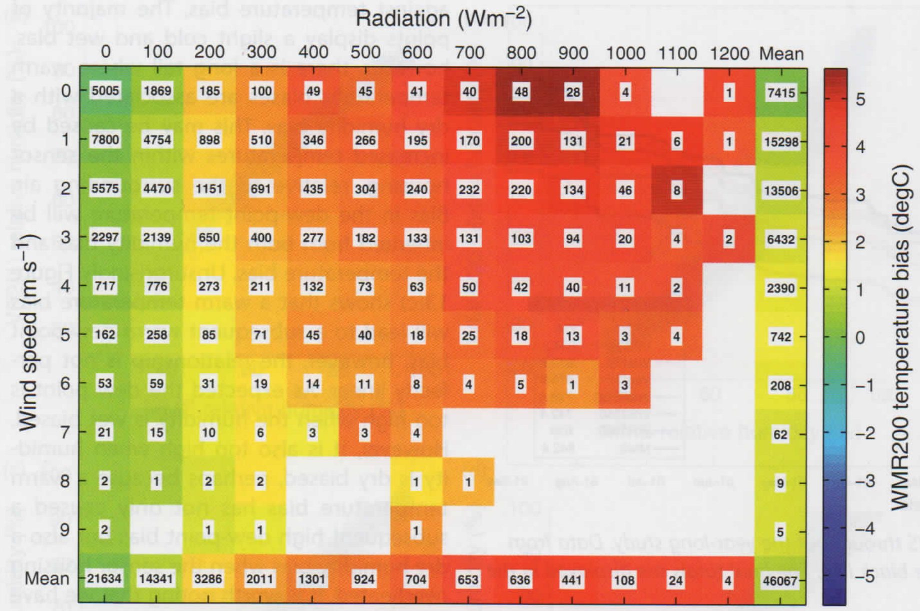

Figure 11. The WMR200 station's temperature bias as function of wind speed and global radiation. The mean bias for a given radiation (wind speed) bin for all wind speeds (radiation levels) is shown along the bottom (right side). Here the number within each cell signifies the sample size.

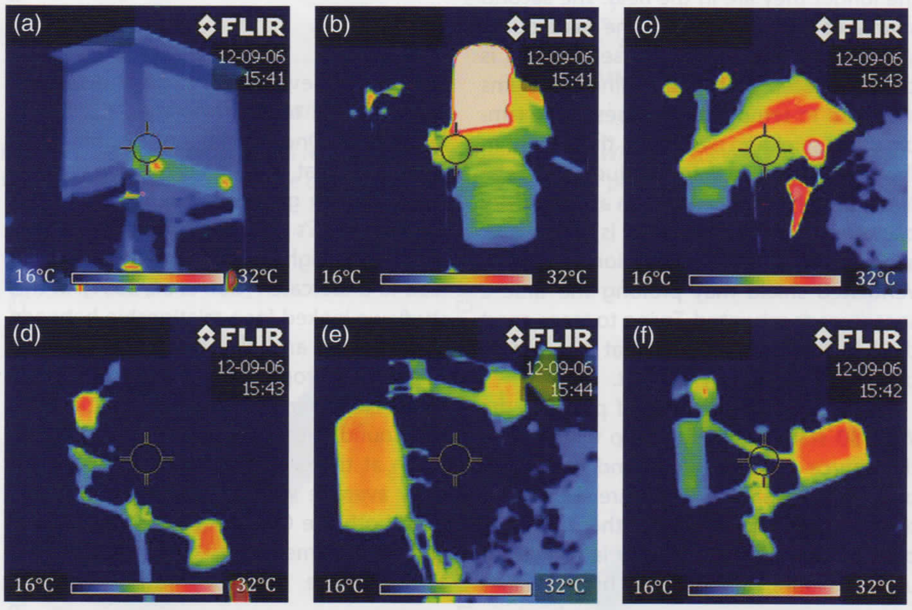

Figure 12. Thermal images taken from the southwest by a Flir i5 thermal imaging camera on a sunny summer afternoon: (a) Stevenson screen and (b) VP2, (c) Vue, (d) WMR200, (e) WS2350 and (f) WH1080 stations. The colour-scale is consistent. The white (hot) part of the VP2 station evident in panel (b) is its black rain gauge. For help identifying the parts of each station, cross-reference with Figure 2, but be aware of the change in perspective.

explanations for this lack of relationship is that the Winterbourne No. 2 site is relatively sheltered, with a mean 7m wind speed of 1.6ms^{-1} and a maximum of 10.2ms^{-1}. The anemometers of the four Davis instruments never read any higher than 5ms^{-1}. It is possible that these relatively low wind speeds did not cause noticeable undercatch and that the biases seen had a different source. The MMS's gauge is deep with steep sides to prevent heavy rain from bouncing out, while all CWS tested, bar the VP2s, are much shallower, particularly the WS2350 and WH1080. Such a design makes them prone to undercatch due to rain drops bouncing out, an effect that potentially outweighs any influence of wind speeds.

Even before deploying the CWS' rain gauges outdoors, the tipping buckets within may be poorly calibrated producing a bias straight out of the box. To test this, 500ml of water was slowly dripped through each rain gauge indoors. By dividing the volume of water by the area of the gauge it is possible to calculate the depth of rain (in mm) that the station's console should display (Overton, 2007). Differences between expected and measured depth were as large as 13%, with most stations underreading. A corresponding correction was then applied to the yearly cumulative rainfall totals. For some stations this led to a small improvement, but for others their yearly total was made much worse. Clearly other factors are at work outdoors that outweigh errors due to poor calibration, proven by the poor performance of the VP2 stations in the field despite being calibrated to less than 2% error indoors.

Some CWS' rainfall bias, particularly when dealing with daily totals, can result from cold and snowy synoptic conditions. The gauges of the CWS tested were prone to filling with snow when the MMS's gauge did not. This resulted in delayed tips when the snow finally melted. The funnel exit hole of the CWS' gauges was also prone to freezing over, again causing a delay in rainfall readings.

Using the results for modelling biases

This experiment was not designed as a stand-alone project – the aim was to aid the construction of models that will estimate the bias and uncertainty associated with CWS' observations. Here we discuss how key findings from the field study influence these models of instrumental bias.

All of the CWS we tested exhibited a relationship between temperature bias and the levels of global solar radiation. The relationship varied depending on the model of station, so ideally we would like to learn this relationship for each individual station type over time. In order to do this we would require an estimate of the solar radiation at the location of all CWS in the country.

Figure 14 demonstrates that with a reliable estimate of the global solar radiation level, as is available at the test site, it is relatively simple to correct the temperature observations. Here we used a simple multiple linear regression model in which radiation, wind speed and an interaction term were used as predictors of the temperature bias, with radiation providing most predictive power. When the temperature bias prediction was used to correct the observations there was a reduction in the mean bias and residual variance for all the CWS. Sometimes this improvement was marginal, for example, for the Davis stations, but for stations that exhibited large radiation biases, such as the WS2350, the improvement was large. As previously mentioned, the influence of radiation on temperature bias is often lagged, so incorporating previous radiation observations using an exponential weighting would improve our correction further.

We also saw that temperature bias can result from poorly calibrated sensors, as

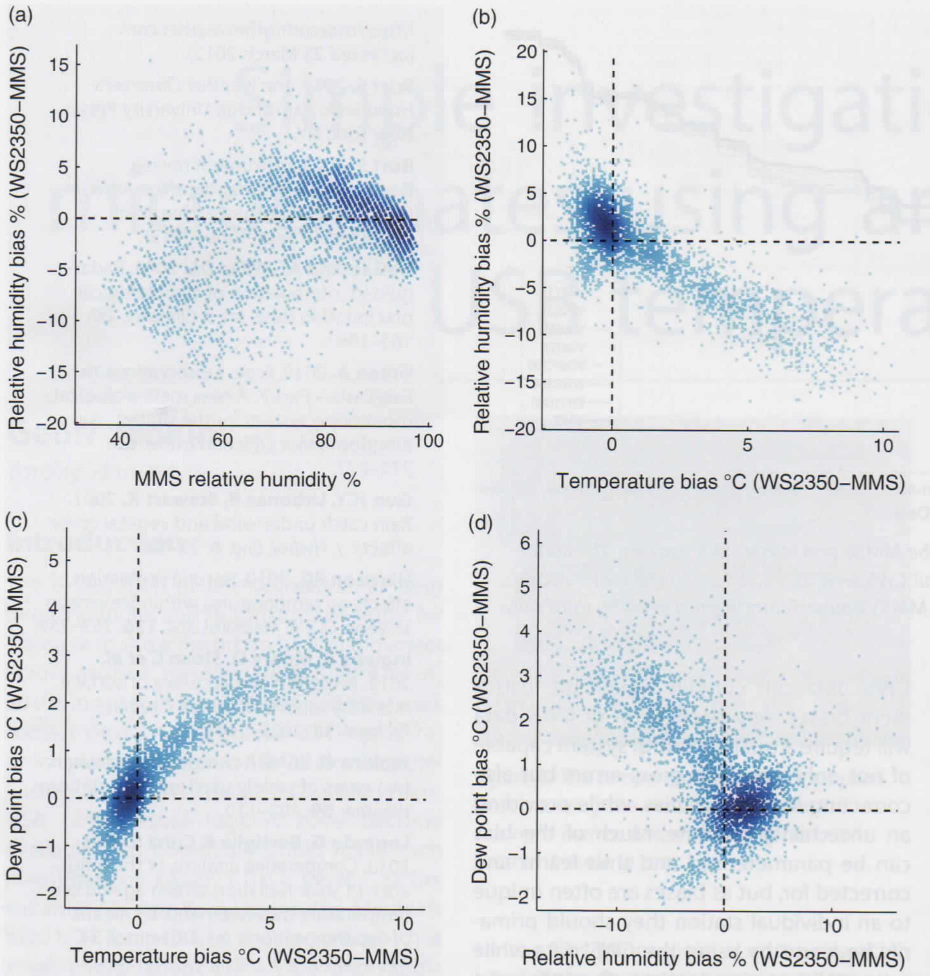

Figure 13. Plots of the relationship between temperature, humidity and dew point, and their biases for the La Crosse WS2350 station. The darker the colour the higher the density of points.

with the Vue(2) station, which displayed a night-time cool bias of around −0.2 degC. We can quantify this calibration bias under conditions when the bias produced by factors such as radiation and wind speed is minimal, for example at night under cloudy, breezy conditions.

The field study has reinforced that relative humidity is a difficult variable to measure, for both CWS and MMS's stations. As the MMS network has its own bias that can slowly drift or suddenly jump, it is clear that this network may have to undergo its own quality control procedures. A network of more accurate sensors, such as chilled mirror hygrometers that measure dew-point temperature directly, could help anchor the MMS network and in turn the CWS network, although such a network may prove difficult to keep in good operational condition. Incorporating relative humidity into our predictive model will also be difficult. With an upper limit capped at 100%, it can produce non-Gaussian errors. Converting relative humidity into other variables that represent air moisture, such as dew-point temperature, may provide a solution to this. We have shown examples where relative-humidity and dew-point biases are dependent on temperature bias, so it is important that we model the moisture variable jointly with temperature and its bias.

The CWS are usually good at capturing the intensity and timing of rainfall events, but their long-term cumulative total can differ from professional gauge measurements significantly. Using the MMS's gauge as our best estimate of the truth, Figure 15 shows that by identifying the relationship between the MMS's and CWS' cumulative rainfall time series we can correct the CWS' time series to fall in line with that of the MMS. This works well because, for the CWS tested here, the proportion of undercatch or overcatch tended to remain relatively constant through time, thereby allowing for a correction to be learnt from preceding data. In reality, a MMS gauge will not be collocated alongside all CWS, in which case observations must be carefully interpolated from nearby professional gauges, for example, from the MMS or Environmental Agency networks. Effectively merging radar accumulation data with this gauge data could improve the interpolation (DeGaetano and Wilks, 2009), which should improve CWS' long-term totals, while keeping the detail of individual and isolated rainfall events captured by the CWS.

We have seen examples where two stations of a particular model exhibit very similar biases, for example the temperature measurements of the two VP2s, although ideally we would have tested at least three

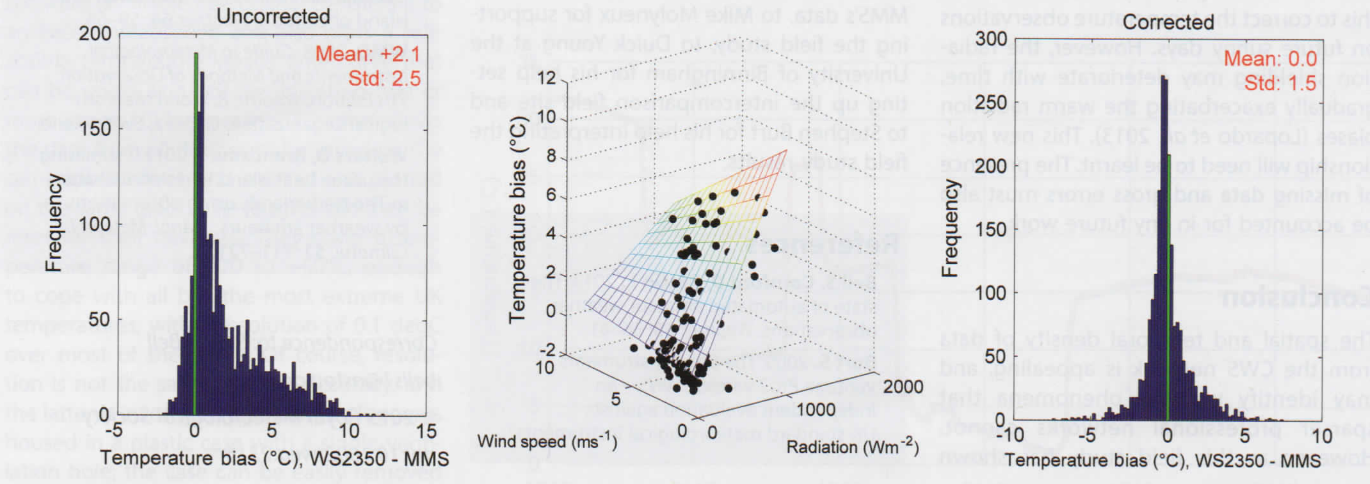

Figure 14. Demonstration of correcting CWS' temperature bias using a multiple linear regression model. The figure shows a histogram of the WS2350 station's temperature bias before and after the correction, along with a scatter plot of a sample of its observations overlaid with a grid of the learnt model. The data were randomly split in half to form the training and test datasets.

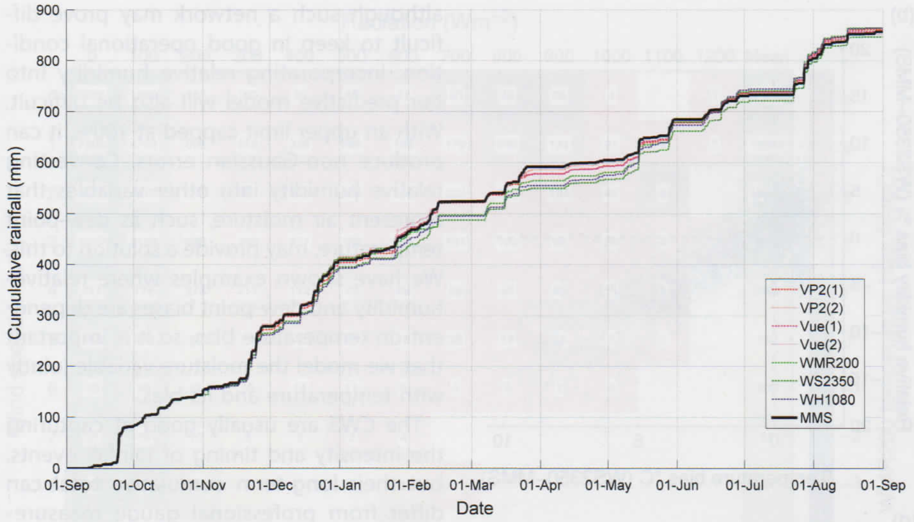

Figure 15. Plot of cumulative rainfall totals from the MMS's and seven CWS' gauges. The cumulative rainfall values (available every 10 min) for all CWS were corrected using the relationship between their cumulative rainfall and that of the MMS's gauge; learnt from preceding data only.

of each model to give higher confidence in the consistency. Station metadata becomes very valuable when it contains details such as station type, providing prior information about the bias we would expect to see. However, as noticed by Jenkins (2014), this study also showed examples when stations of the same model actually displayed very different biases (e.g. the VP2 stations' rainfall or the Vue(2)'s night-time cold bias not evident in the Vue(1)). Within the full CWS network all CWS will have a different siting, exposure and level of calibration and maintenance. Each can introduce biases unique to a particular site. For these reasons our study is only representative of the biases in the network up to a point.

This study has shown examples of how we can correct the bias in one variable by evaluating its relationship with other variables. It is important that these relationships remain smooth through time while still being able to adapt and update. If, for example, we learn that a station has a significant warm bias under sunny conditions, we can use this to correct the temperature observations on future sunny days. However, the radiation shielding may deteriorate with time, gradually exacerbating the warm radiation biases (Lopardo *et al*., 2013). This new relationship will need to be learnt. The presence of missing data and gross errors must also be accounted for in any future work.

Conclusion

The spatial and temporal density of data from the CWS network is appealing, and may identify weather phenomena that sparser professional networks cannot. However, as this field study has shown CWS' data can contain significant instrument biases. Any application of CWS' data will require a quality control system capable of not only removing gross errors but also correcting instrument bias, while providing an uncertainty estimate. Much of the bias can be parameterised and thus learnt and corrected for, but as biases are often unique to an individual station they should primarily be learnt by using the CWS' data while potentially varying in time. Quantifying a station's bias will rely on obtaining a reliable estimate of the weather at CWS' locations against which we can begin to verify the CWS' observations, and this is the subject of our future work. This will attempt to disentangle instrument biases from the natural spatial variations that we wish to capture.

Acknowledgements

This research is funded by an EPSRC CASE award (10002388) with the Met Office. Thanks are due to Shona Hogg and Fiona Carse at the Met Office for providing the MMS's data, to Mike Molyneux for supporting the field study, to Duick Young at the University of Birmingham for his help setting up the intercomparison field site and to Stephen Burt for his help interpreting the field study results.

References

Bell S, Cornford D, Bastin L. 2013. The state of automated amateur weather observations. *Weather* **68**: 36–41.

Burt S. 2009. The Davis Instruments Vantage Pro2 wireless AWS – an independent evaluation against UK-standard meteorological instruments. http://measuringtheweather.com (accessed 23 March 2012).

Burt S. 2012. *The Weather Observer's Handbook*. Cambridge University Press: New York, NY.

Burt S. 2013. Instrument review Davis Instruments Vantage Vue AWS. http://measuringtheweather.com (accessed 16 September 2013).

DeGaetano AT, Wilks DS. 2009. Radar-guided interpolation of climatological precipitation data. *Int. J. Climatol.* **29**: 185–196.

Green A. 2010. From Observations to Forecasts – Part 7. A new meteorological monitoring system for the United Kingdom's Met Office. *Weather* **65**: 272–277.

Guo JCY, Urbonas B, Stewart K. 2001. Rain catch under wind and vegetal cover effects. *J. Hydrol. Eng.* **6**: 29–33.

Harrison RG. 2010. Natural ventilation effects on temperatures within Stevenson screens. *Q. J. R. Meteorol. Soc.* **136**: 253–259.

Ingleby B, Moore D, Sloan C *et al.* 2013. Evolution and accuracy of surface humidity reports. *J. Atmos. Oceanic Technol.* **30**: 2025–2043.

Jenkins G. 2014. A comparison between two types of widely used weather stations. *Weather* **69**: 105–110.

Lopardo G, Bertiglia F, Curci S *et al.* 2013. Comparative analysis of the influence of solar radiation screen ageing on temperature measurements by means of weather stations. *Int. J. Climatol.* **34**: 1297–1310.

Muller CL. 2013. Mapping snow depth across the West Midlands using social-media-generated data. *Weather* **68**: 82.

Overton AK. 2007. A guide to the siting, exposure and calibration of automatic weather stations for synoptic and climatological observations. http://myweb.tiscali.co.uk/awsguide (accessed 10 October 2013).

Steeneveld GJ, Koopmans S, Heusinkveld BG *et al.* 2011. Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in the Netherlands. *J. Geophys. Res.* **116**: D20129.

Wilby RL, Jones PD, Lister DH. 2011. Decadal variations in the nocturnal heat island of London. *Weather* **66**: 59–64.

WMO. 2008. *Guide to Meteorological Instruments and Methods of Observation*, 7th Edition, Volume 8. World Meteorological Organization: Geneva, Switzerland.

Wolters D, Brandsma T. 2012. Estimating the urban heat island in residential areas in The Netherlands using observations by weather amateurs. *J. Appl. Meteorol. Climatol.* **51**: 711–721.

Correspondence to: Simon Bell

bells1@aston.ac.uk

doi:10.1002/wea.2316