Report on calibration of the EODG Weather Station

John Evans
Supervisor: Dr. Adam Povey

September 26, 2014

Contents

1 The Weather Station

The Earth Observation Data Group, part of the Department of Physics at the University of Oxford, operates an automated weather station on the roof of the Atmospheric Physics building (51.75927°N, 1.25568°W, altitude 60m AMSL) , providing information on the current conditions on an automatically updated webpage, and also raw measurements dating back several years. Although the unit, which is a Met4Net station supplied by Instromet Weather Systems Ltd, was factory calibrated, the accuracy and consistency of data output has not previously been investigated. As part of a vacation student project several comparisons to other data sources have been completed in an attempt to quantify any systematic errors and miscalibrations, and to produce a recommended offset or calibration function for each sensor if necessary. Data exists continuously from the present time back until August 2012, when the current weather station was activated. Data from the period 2001-2008 also exists from a previous station on this site, but as this was not the same model of station this data was not considered further in the calibration.

1.1 Data Collection

The Met4Net system includes a dry bulb thermometer, a relative humidity sensor, a raingauge, wind direction, wind speed, and barometric pressure sensors. The sensors on the roof are attached to a datalogging unit in the office below, which can operate independently but in this case is permanently linked to a computer. All but the pressure sensor are located on the roof; the pressure sensor is inside the datalogger itself. The datalogger unit reads the instantaneous sensor values every 0.5 seconds, and internally averages them over a 10 second period before writing them to internal memory. Every 2 minutes the controlling computer reads these values out and writes them to a plaintext file on disk, and updates the webpage with the latest data. Every 9am GMT the previous 24 hours of data are uploaded to the online repository.

1.2 Calibration Method

As an initial check on the data consistency and correct operation of the EODG Station, the data log was compared to a nearby separate source of weather data available for as much of the station's period of operation as possible. It would be preferable to compare to another automated weather station taking readings continuously, so that the entire time series could be compared directly, but there does not appear to be one nearby. The nearest provider of meteorological data is the Radcliffe Observatory, which is approximately 600m from the EODG Station, and has roughly the same sort of exposure and environment ie. Central Oxford, and so we will assume that the same meteorological values should be measured at both locations at any one time. The values from the Radcliffe Observatory were assumed to be the 'true' meteorological values.

The aim is to produce a set of "calibration coefficients" that can be applied to the raw data coming out of the weather station; for example, by plotting the "true" values from the comparison dataset on the x axis and the equivalent EODG values on the y axis, then finding a fit to the data. The inverse of the function found will then give a calibration function that could be applied to data from the EODG Station to improve the quality of the data.

2 Comparison to the Radcliffe Observatory

The Radcliffe Observatory, also known as the Radcliffe Meteorological Station, is situated in Green Templeton College Oxford (51.76083°N, 1.26385°W, altitude 63.4m AMSL) and is the site of the longest series of temperature and rainfall records for one site in Britain. The observatory was built between 1772 and 1774 for astronomical observation and air temperature measurements were taken to support this use. The original Radcliffe Observer, Dr Thomas Hornsby, was also interested in the weather and so sporadically took other observations. However, regular reporting of meteorological observations did not begin until the middle of the 19th century. In 1935 the work of the Observatory was moved to South Africa, but observations have continued to this day.

Since 1925 observations at the Observatory have consisted of a single eye-observation made at 0900 GMT every day, consisting of spot measurements and reporting of data from the past 24 hours. Additionally, wind speed and direction data is available on an hourly basis, as the wind data is taken from an anemometer mounted to the top of the Engineering Department.

2.1 Data Provided

The Radcliffe Observatory provided 9am observation data in monthly spreadsheets from August 2012 to May 2014. Some of the data provided, such as temperature above grass and concrete, soil temperatures, observed visibility, and snow depth, is not taken by the EODG Station, and so did not form part of the comparison. Those parameters that were compared were dry bulb temperature, barometric pressure, relative humidity and windspeed - all taken at 0900 GMT - as well as the accumulated rainfall, hours of sunshine, maximum temperature and minimum temperature in the 24 hour period.

Typical meteorological practice is to record these accumulated or max/min values at the observation time and reset the recorders at that point; ie. 24 hourly rainfall recorded at 0900 GMT on the 26th June is the total rainfall recorded between 0900 on the 25th June and 0900 26th June. However, the Radcliffe Observatory does not follow this practice with respect to the maximum temperature, rainfall and sunshine hours accumulated - instead they publish the values recorded in the 24 hours *after* the observation time for these readings. This is presumably because then the values make more sense intuitively - for example maximum temperatures are most likely to happen in the middle of the day, and so if recorded correctly a high temperature at midday climate and on the 25th would appear on the record against the 26th, as that would be when the maximum thermometer would be read and reset.

Although correct this is confusing for someone who is casually browsing the data, as the maximum temperature appears to be against the "wrong" day; so the Observatory shifts these three readings by a day to make them appear on the intuitively correct day. This initially caused some confusion and it should be noted that the Radcliffe Observatory website explicitly and incorrectly states that all data is for the previous 24 hours.

2.1.1 Data Processing

The Radcliffe data was condensed into a single log file for the complete time period to make it easier to process, and days with missing data were deleted. Additionally, days with any comments against them in the Radcliffe source data noting data collection problems for that day eg. a frozen or dry wet bulb, broken equipment, or obvious human error, were excluded from the comparison. Any days with missing data from the EODG Station were also deleted, although as the station is automated this only amounted to a handful of occasions where the power had failed. This deletions resulted in a decrease in the amount of data available for comparison but it was not significant when compared to the number of days in total - after this process there were 547 days with valid data to compare. Additionally, where rainfall had been recorded as "Trace" ie. some precipitation had fallen but less than the first graduation on the recorder, this was assumed to be zero, as due to the automated nature of the EODG Station it would record nothing in this scenario.

A small C++ script was written to pull together both this data from the Radcliffe Observatory, and also to fetch the appropriate data from the repository of EODG Station data. The first logged observation after 0900 GMT was used as a comparison point, which is always within 10 seconds of the hour(the exact time depends on the drift of the internal clock of the datalogger, which is resynchronised to global time every day.) For values requiring cumulative values over 24 hours the script scanned to 0900 GMT the previous and next days and found the maximum and minimum temperatures, as well as the accumulated rainfall and sun hours over these time periods as described above. Both sets of data were then written out to a single file for analysis using Gnuplot.

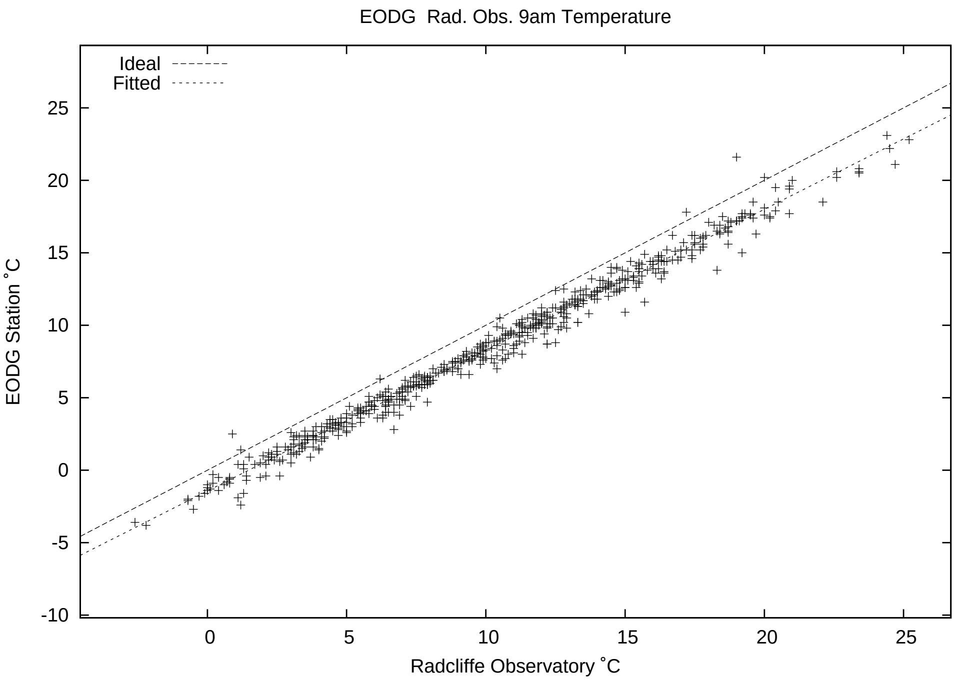

Figure 1: Comparison of 0900 temperature readings

2.2 Results

2.2.1 Temperature

Figure 1 shows the 0900 temperature readings for the Radcliffe Observatory and the EODG station plotted against each other. A line is drawn where the ideal 1:1 fit would be, as well as the linear fit line to the data. A linear fit $(\mathrm{a} + \mathrm{b}x)$was chosen as it could be reasonably expected that any non-linearities in the EODG temperature sensor had been calibrated at the factory, leaving only some offset to the true value, and visually the plot does not seem to exhibit any non-linearities, although at the top and bottom ends there is less data and so it is a little more difficult to tell. The constant term in the fit evaluated to $\mathrm{a} = -1.43614 \pm 0.05907$, and the linear term $\mathrm{b} = 0.971807 \pm 0.005137$ using a least squares algorithm. These fit parameters effectively give calibration factors for the EODG station. The reduced chi-square value for this fit is 0.45089, however this number is not very useful as we put no errors or weighting on the data points(the lack of uncertainties will be returned to later.) We can still easily see both visually and from the standard deviation on the fit parameters that the fit seems representative of the data though.

Similarly figure 2 shows the maximum temperature over 24 hours plotted against each other. This was also fitted using a linear fit and the parameters given for the fit were $\mathrm{a} = -1.32781 \pm 0.07428$ for the constant term, and for the linear term $\mathrm{b} = 1.00636 \pm 0.004964$. These values are similar to those given for the previous temperature fit, which is reassuring. We can see that the values are still well correlated, suggesting that other than a constant offset due to calibration the exposure and environment of the two temperature sensors is comparable enough to produce a similar maximum temperature on a given day.

Figure 3 on the other hand shows the minimum temperature over 24 hours for each of the comparison days. Although clearly still positively correlated the correlation is not quite as strong, suggesting that there are differences in the sensor exposure overnight that can introduce more of a difference into the minimum temperature; the EODG station is mounted above a roof which is covered in roofing felt, whereas the Radcliffe Observatory temperature sensor is mounted in a grass covered quad. This could mean that the surfaces retain different amounts of heat overnight and create a different thermal environment over the surface. However, this difference is not dramatic, the fit parameters ($\mathrm{a} = -1.02384 \pm 0.05594$ for the constant term, $\mathrm{b} = 0.978514 \pm 0.007103$ for the linear) do not exhibit large fit uncertainty and are

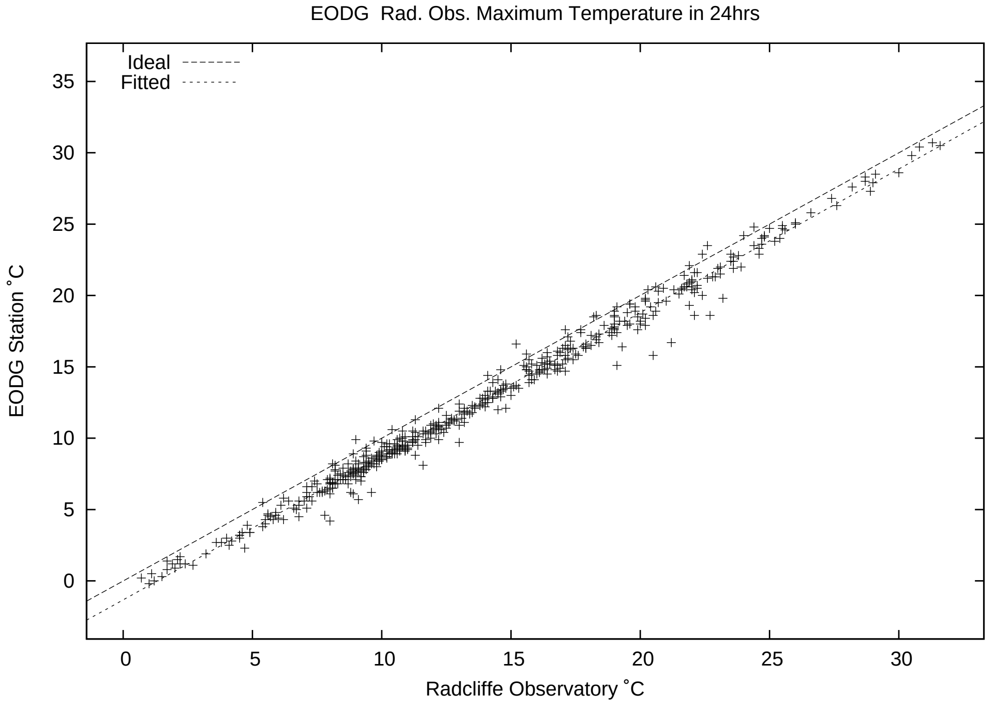

Figure 2: Comparison of daily maximum temperatures

again similar to that obtained in the other two temperature fits.

2.2.2 Pressure

Figure 4 shows the comparison of the sea level pressure readings at 0900. When comparing pressures between two different locations it is important to allow for any difference in altitude, as atmospheric pressure decreases with increasing height irrespective of the weather. It is possible to use the barometric formula:

$$p_0 = pe^{\frac{gMh}{RT}}$$

to convert to a sea level pressure (p) given a pressure at altitude h (p_0) and using the surface gravitational acceleration (g), molar mass of dry air (M), universal gas constant (R) and the temperature T.

Having converted both sets of pressure readings to be sea level pressures it is therefore possible to directly compare them to each other. We can see that there is generally very good correlation between the two values, although the slope of the fit is a little way off being the ideal 1:1. There is also some noise apparent where one of the weather stations has read a dramatically different value to the other; from this data alone it is not possible to determine which one was correct on any given day. The fit parameters given were a $= 123.463 \pm 14.52$ for the constant term, b $= 0.8818 \pm 0.01433$ for the linear.

For this particular sensor only the first 95 days were ignored in the comparison, as a large unknown offset had been accidentally applied to the EODG sensor data inside the datalogging software; the reduction in data available does not seem to have affected the validity of the comparison.

2.2.3 Relative Humidity

Figure 5 plots the relative humidity readings from each location at 0900. Immediately it is obvious that this plot is much nosier and less well correlated, although the values do appear to mainly track the ideal 1:1 line. There are a few values greater than 100

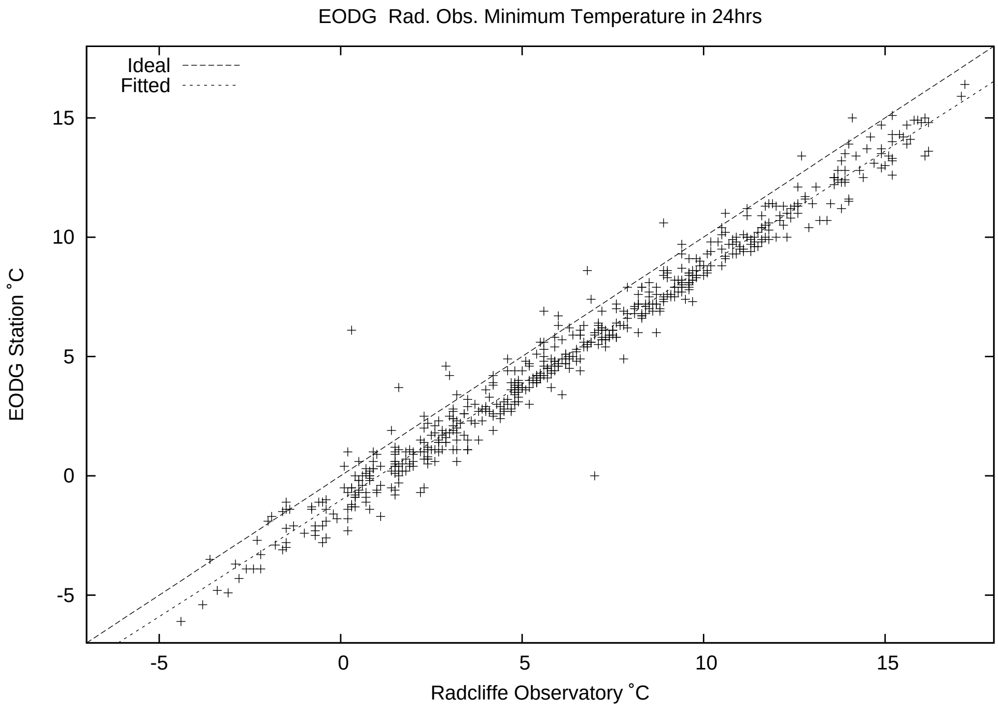

Figure 3: Comparison of daily minimum temperatures

However there is also a curious feature at the top end of the range where the graph appears to 'turn off' and there is a set of values where the Radcliffe Observatory has recorded high values of humidity where the EODG Station did not. Although again it is impossible to tell which measurement is in error from this data alone, the difference could be explained by the different types of measurement technique involved; the EODG Station uses a solid state electronic humidity sensor but at the Radcliffe Observatory values are calculated using the separately measured wet and dry bulb temperatures. Human error while reading one of the temperatures, or the wet bulb not being completely wet, could explain the occasions on which this happens.

No fit was attempted to this data as without confirmation from another data source it is difficult to tell if there is a problem with the EODG data or the Radcliffe Observatory data here.

2.2.4 Windspeed

Figure 6 shows the windspeed readings at 0900. Data from both stations exhibits a large step size , but a larger problem is that the windspeed read at EODG always seems to be much lower than that read at the Radcliffe Observatory. There are probably two main reasons for this: unlike, say, pressure, windspeed can vary by an order of magnitude across only a few seconds or across lengthscales of hundreds of meters. This reading is taken at a single point in time and so it is unlikely that the windspeed is the same in each location at 9 am. Additionally, the EODG wind sensor, while pole mounted, is only about 3m above roof level in an area with several buildings, so is probably still in the turbulent flow region close to the ground, whereas the Radcliffe Observatory uses data from an anemometer mounted on a larger pole on top of a small tower block that stands above the surrounding buildings, so it is likely into the laminar flow region where the windspeed is generally steadier and higher. The wind sensors are within sight of each other, and anecdotally, standing and observing both sensors seems to suggest that the Radcliffe Observatory anemometer spins much more regularly than the EODG sensor, so even if instead the average windspeed over a period of an hour (for example) was compared it would not necessarily be the same. Although the graph suggests weak positive correlation, due to these reasons it is difficult to gain any useful information from this data to suggest if any sort of calibration is needed for the EODG

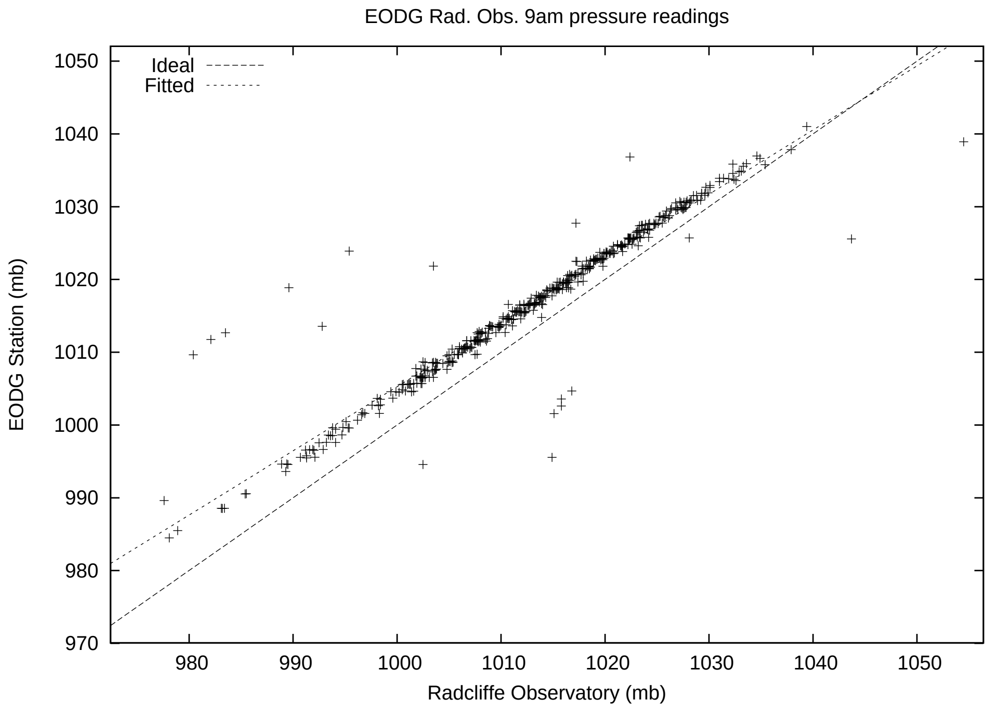

Figure 4: Comparison of 0900 pressure readings

wind sensor.

2.2.5 Sunshine hours

The number of hours of sunshine per day recorded by each station are plotted on figure 7. Although there is positive correlation there is a very wide spread of values about the trend line; generally it seems that the EODG station records less hours than the Radcliffe Observatory, but not always. The EODG sunshine sensor is a light sensor which records the presence of the sun while the output voltage is over a certain factory calibrated value; and it is assumed the Radcliffe's sensor is similar. Any mismatch in the levels between the two would result in an offset so that if the sun is only weak (eg. during winter) then only one station would pick up the sun's presence, leading one station to record zero hours of sun. However, the graph appears to exhibit both stations doing this, as near the 0-0 point we can see two heavy clusters of data points, some where the EODG station has recorded nothing but the Radcliffe has a non-zero value, and vice-versa. This is odd as we would expect that if one station has a different recording level to the other only one of these features would exist.

A tentative fit was applied to the data and the results were a $= 0.110767 \pm 0.09836$ for the constant term, b $= 0.741107 \pm 0.01726$ for the linear.

2.2.6 Rainfall

The rainfall per day, on the other hand, shown in figure 8, is very well correlated and appears to almost exactly agree between the two stations. There is a small amount of noise, which could be due to occasional faults such as the rainfall inlet port being blocked temporarily, but overall this plot is very encouraging. The fit parameters are a $= 0.0762897 \pm 0.0444$ for the constant term, b $= 1.00209 \pm 0.008999$ for the linear.

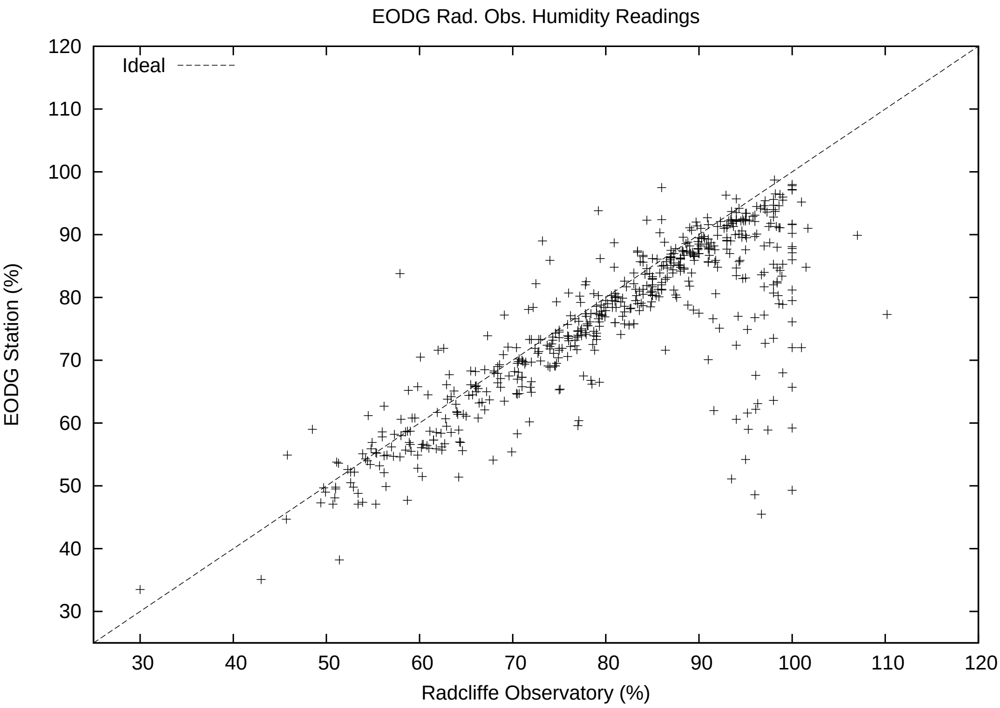

Figure 5: Comparison of 0900 relative humidity readings

2.3 Problems with this analysis

Although several of the plots here appear to confirm that the EODG Station and the Radcliffe Observatory correlate well, particularly temperature, pressure, and rainfall, from which we can get appropriate calibration offsets; several of the plots exhibit problems, such as the odd shape of the relative humidity plot. Additionally, although we assumed that the Radcliffe Observatory data gave the ”true” values at any given if there are problems it is difficult to impossible to tell which data source is incorrect, as we have no comparison. Therefore, in order to try to resolve some of these issues it would be prudent to compare the EODG Station to another data source. Unfortunately no comparable data source other than the Radcliffe Observatory appears to exist in central Oxford, so we will have to look further afield.

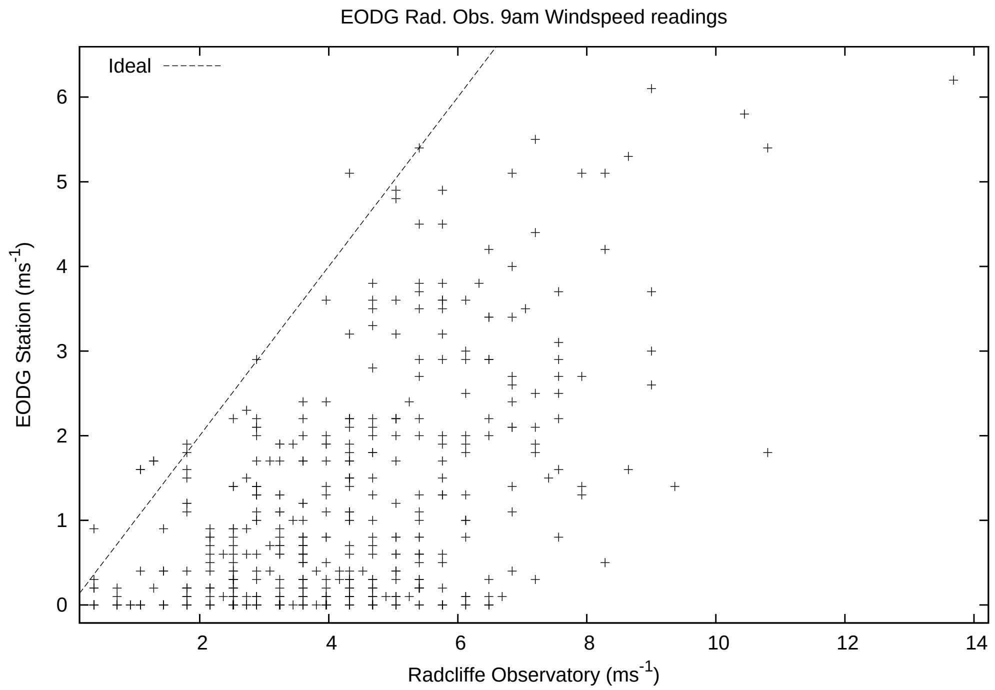

Figure 6: Comparison of 0900 wind speed readings

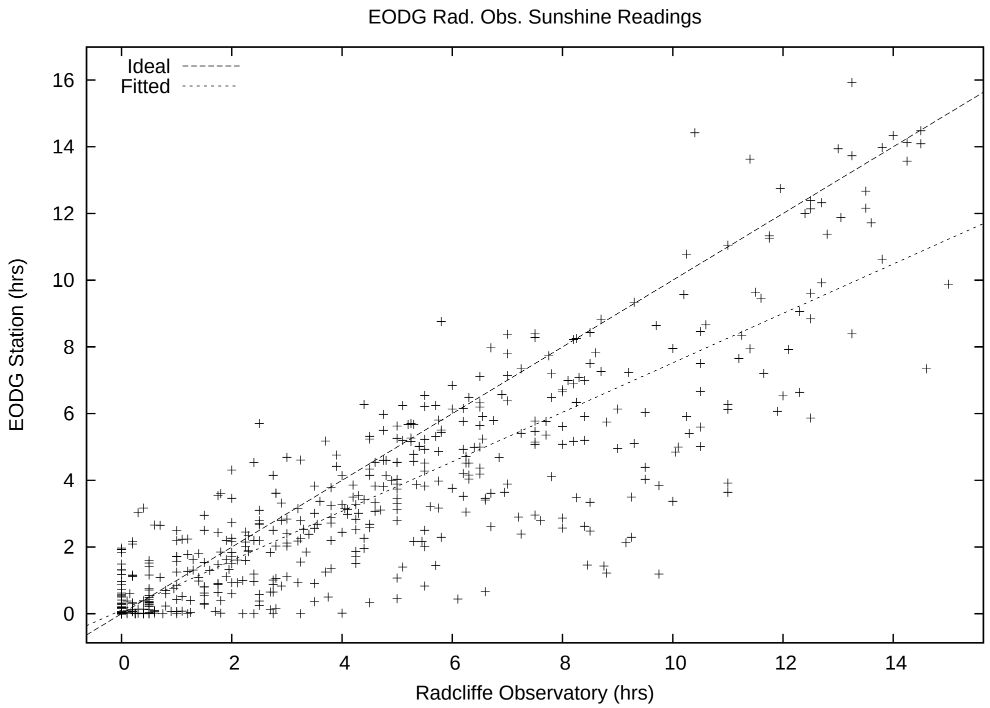

Figure 7: Comparison of 24 hour sunshine readings

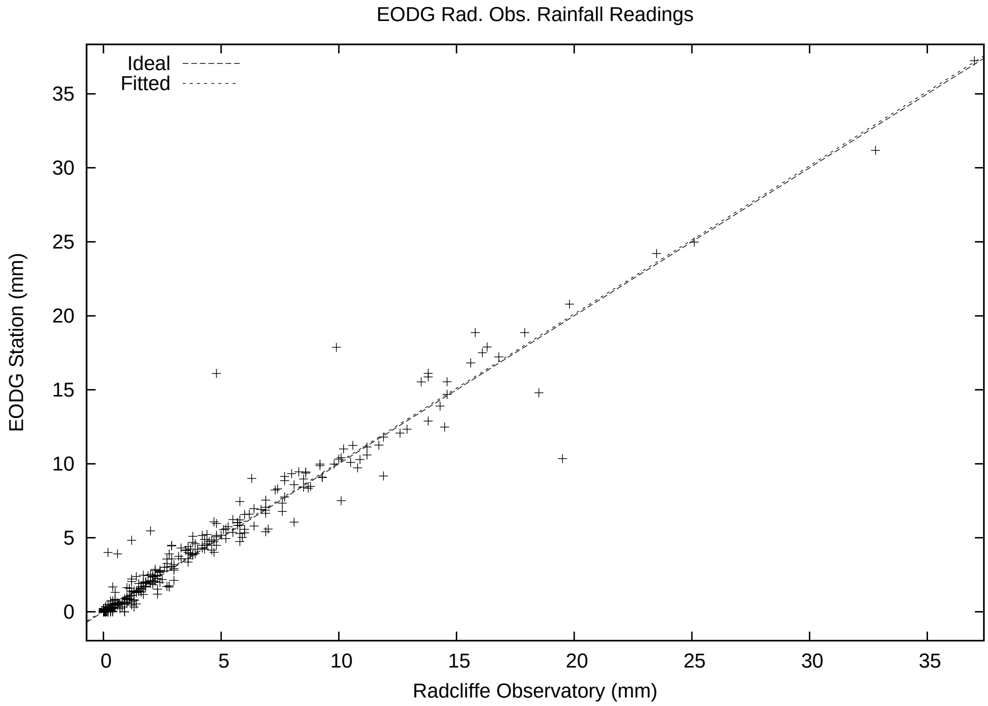

Figure 8: Comparison of 24 hour rainfall readings

3 Comparison to RAF Brize Norton

3.1 MIDAS

To find the nearest comparable source of weather observations the Met Office Integrated Data Archive System (MIDAS) was consulted, an online database containing observations from tens of thousands of land and marine surface stations from 1853 onwards. This extensive dataset would be quite difficult to deal with in raw form but fortunately there are online tools available on the British Atmospheric Data Centre (BADC) website to extract data and export it into easily readable formats. Upon searching for observations for the period of interest closest to Oxford city centre the closest data source (apart from the Radcliffe Observatory) is RAF Brize Norton, an airfield located approximately 16km to the west of the city. Another set of comparisons was therefore conducted with the same dataset from the EODG Station as before in order to double-check the results from the Radcliffe Observatory. Of course, comparing to a station 16km has limitations; values that change on small lengthscales like windspeed are not likely to be directly comparable, but values such as temperature and pressure that change on longer lengthscales should provide useful data. Additionally, the differing exposures of open airfield as compared to city centre may complicate the comparison. Unlike the Radcliffe Observatory, the station at Brize Norton is an automated weather station, like the EODG station, and so can provide data over shorter time periods than 24 hours; most data is given on an hourly basis. However, it was decided to compare data in the same format as the Radcliffe Observatory ie. compare spot readings at 0900 GMT and compare the cumulative sunshine and rainfall readings as well as the maximum and minimum temperatures over a 24 hour period.

3.2 Results

3.2.1 Temperature

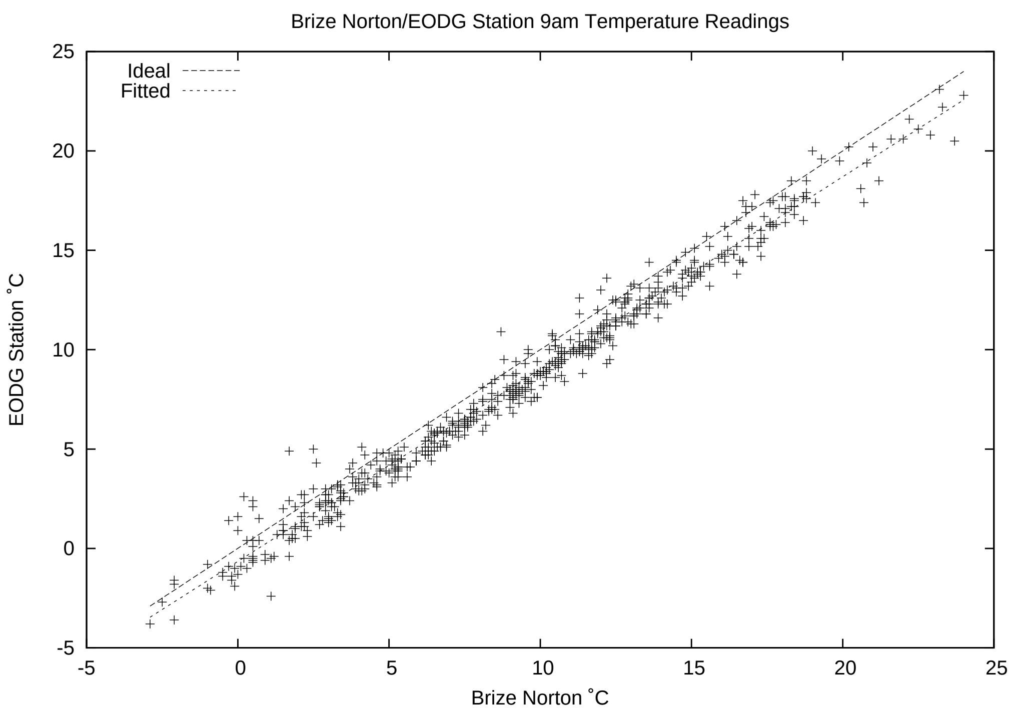

Figure 9: Comparison of 0900 temperature readings

Figure 9 shows the temperature readings at 0900 at Brize Norton and the EODG station. On comparing this graph back to figure 1 which was the equivalent comparison to the Radcliffe Observatory

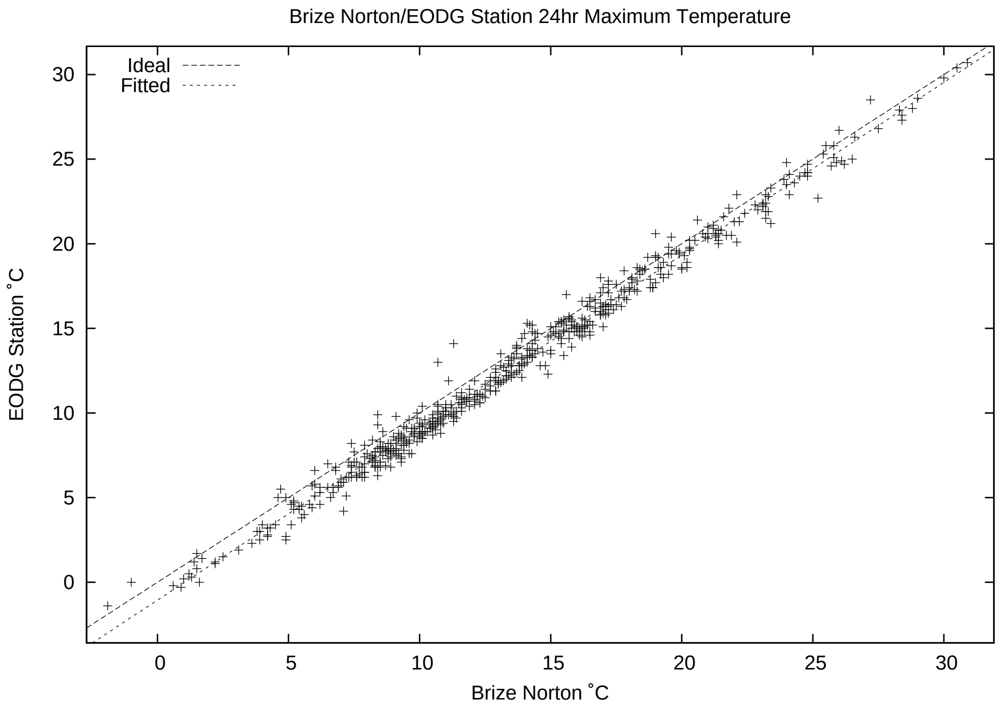

Figure 10: Comparison of daily maximum temperatures

these graphs are encouragingly similar - the correlation is equally good despite the differences in location and exposure. A linear fit $a + bx$ was again used and the fit parameters output were $a = -0.654272 \pm 0.06487$, and $b = 0.967576 \pm 0.005988$. The linear offset here in particular is a little different to the previously obtained value but this comparison at least provides a sanity check on both the EODG and Radcliffe Observatory data.

The comparison of maximum temperatures over 24 hours is in figure 10. The fit parameters for the linear fit are $a = -1.05663 \pm 0.07351$, and $b = 1.01976 \pm 0.005033$, while the comparison of minimum temperatures over 24 hours is in figure 11. The fit parameters for this linear fit are $a - 0.182486 \pm 0.06982$, and $b = 0.902808 \pm 0.009007$. While these fits, particularly the minimum, do not correlate quite as well as the Radcliffe Observatory, as should be expected due to the different environment, they add additional confidence to the EODG data.

Overall the temperature fits seem to confirm that the data being gathered by the EODG Station is of good quality and correlates well with other sources; the small differences in linear offsets between the different stations can be most probably attributed to the different exposure and location.

3.2.2 Pressure

Figure 12 shows the comparison of 0900 sea level atmospheric pressure readings between Brize and the EODG Station, with both sets of readings compensated for altitude. The correlation between the two is exceptional, with no noise apparent at all - due to the large lengthscale that pressure varies over the difference in location between Brize Norton and the EODG Station is mostly irrelevant for this comparison. Additionally, the linear fit parameters, $a = 126.462 \pm 13.9$, and $b = 0.871405 \pm 0.01373$, are almost the same as those obtained for the comparison to the Radcliffe Observatory, adding additional confidence to these calibration parameters.

3.2.3 Relative Humidity

Although the correlation for the relative humidity comparison is not as tight as the Radcliffe Observatory(again likely due to the difference in location being enough to change the humidity/precipitation

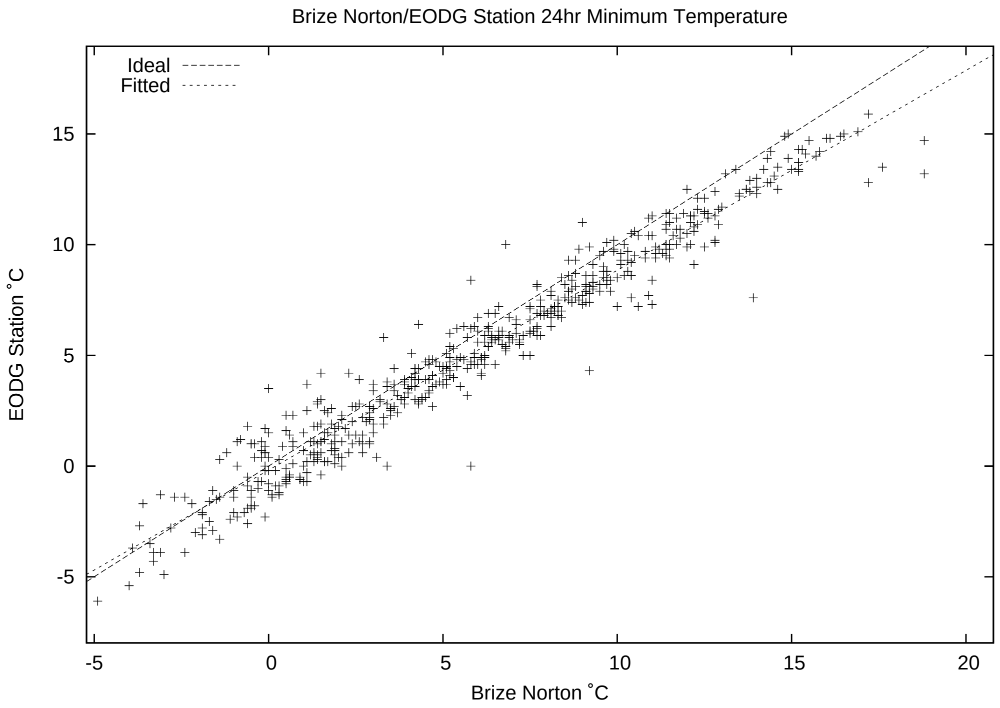

Figure 11: Comparison of daily minimum temperatures

significantly), the graph, shown in figure 13, does not exhibit the unusual turn off at high humidity values, appearing to confirm that this was an artifact of the Radcliffe Observatory data rather than the EODG Station. The linear fit parameters are a $= 2.34902 \pm 1.36$, and b $= 0.960655 \pm 0.01652$, and do suggest a slight underestimation of the true value by the EODG humidity sensor.

3.2.4 Windspeed

Unfortunately, the windspeed comparison in figure 14 comes out quite similarly to the previous windspeed comparison with a very indistinct uncorrelated set of data points, with the EODG windspeed once again much lower than the comparison at almost all points, suggesting that there is either a large problem with the sensor or the local exposure is such that the windspeed is greatly reduced at the sensor, which is a distinct possibility due to the low pole height of the EODG sensor when compared to the surrounding buildings.

3.2.5 Sunshine

Figure 15 shows the comparison of the 24 hourly sunshine hour readings - this shows definite positive correlation although the linear fit, and also the presence of a feature at the bottom of the graph where there are a number of values where Brize has recorded a non-zero value and the EODG Station has recorded nothing, both suggest that the EODG sensor under reads the true number of hours of sunshine. The fit parameters are a $= -0.0807917 \pm 0.06483$, and b $= 0.910935 \pm 0.01308$.

3.2.6 Rainfall

Finally figure 16 shows the comparison of 24 hour rainfall readings between Brize and EODG; this data is not nearly as good as the equivalent for the Radcliffe, undoubtedly because of the different locations. There is still positive correlation however there is not really enough data at higher rainfall levels to give a confidant fit, as most of the data is clustered at the low end. However, the Radcliffe comparison for this data was very encouraging and so this does not seem to be a problem.

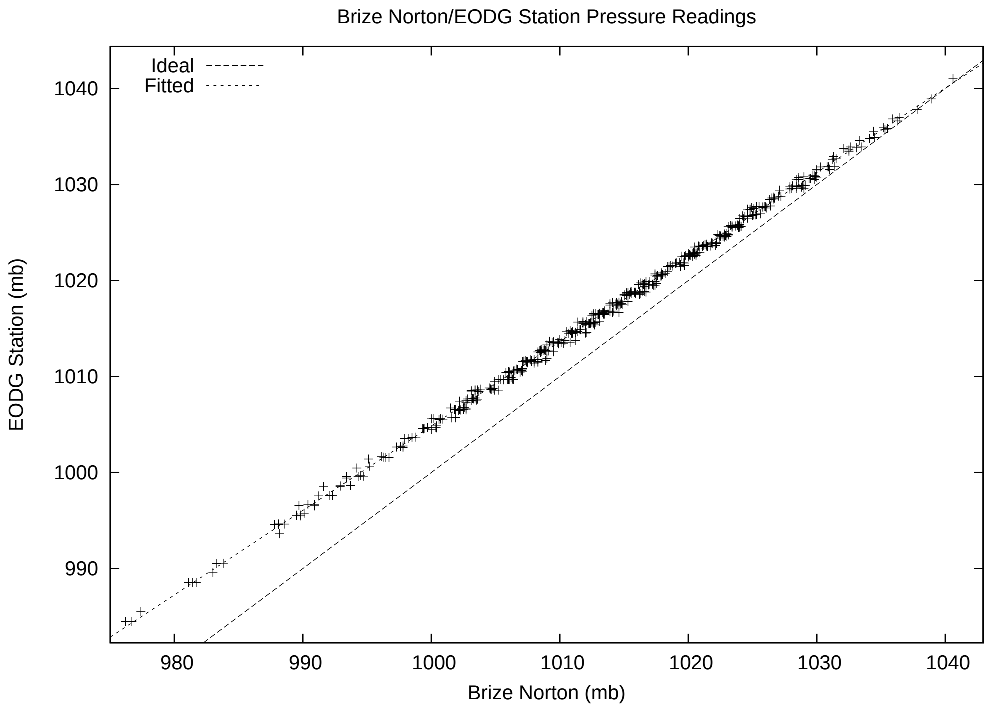

Figure 12: Comparison of 0900 pressure readings

4 Summary of results from these comparisons

Having compared the EODG Station to two different data sources we can see that we can confidently declare the majority of the station output data seems to be of good quality and comparable to data from nearby sensors. In particular comparing pressure, the easiest variable to compare over distance, we get two sets of extremely similar calibration coefficients, so we can be confident that these are good offsets to apply to the data. Temperature is a little more difficult as we have obtained several different values particularly for the intercept parameter - some of this is probably due to the different environment of the different sensor locations giving different thermal environments.

The sunshine sensor is difficult to calibrate as due to the design of the sensor it is only an "on-off" reading, so simply applying a linear offset to the output is not enough to correctly calibrate - for instance, imagine a scenario where the sun is just over the level required to count as sunshine as defined by the Met Office but is just below the sensors trigger value all day - zero sun hours will be recorded, just the same as if the radiation flux was below the correct trigger level, and so it is not currently possible to differentiate between these possibilities. Some sort of continuously variable light sensor would be much more helpful for calibration.

The windspeed has proved the most difficult to compare - it is very locally variable in both time and space and so comparing a single moment in time in two different locations does not seem to work very well at all. It might be possible to instead compare the average windspeed over an hour but this does not account for differences in local exposure. Instead it would be best to compare to another collocated anemometer that has been correctly calibrated in a wind tunnel so that there was no difference in location.

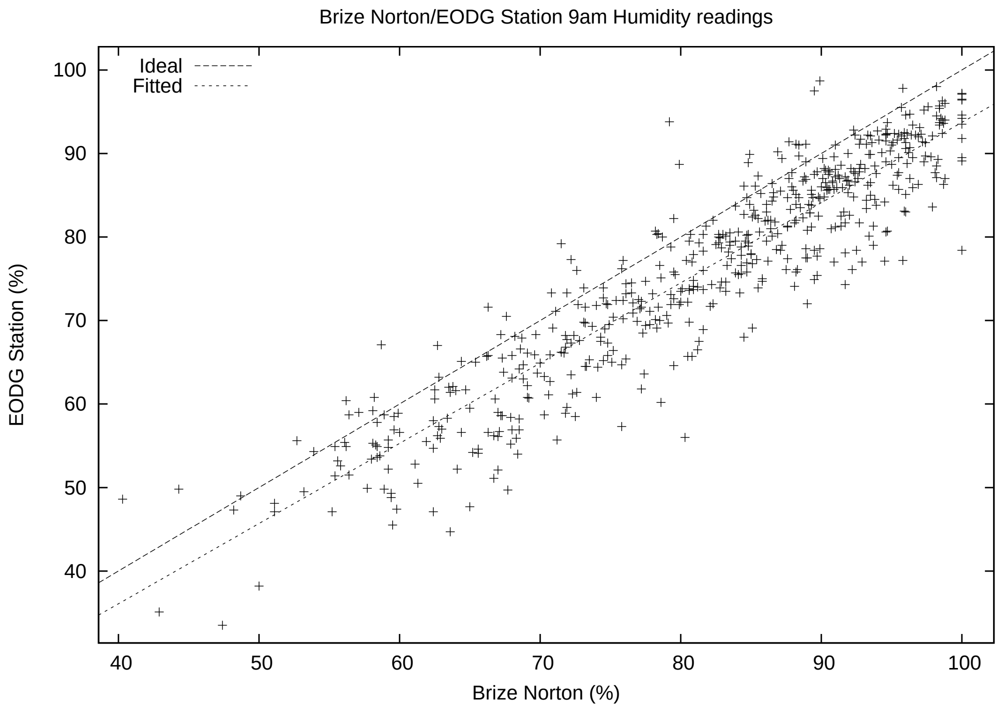

Figure 13: Comparison of 0900 relative humidity readings

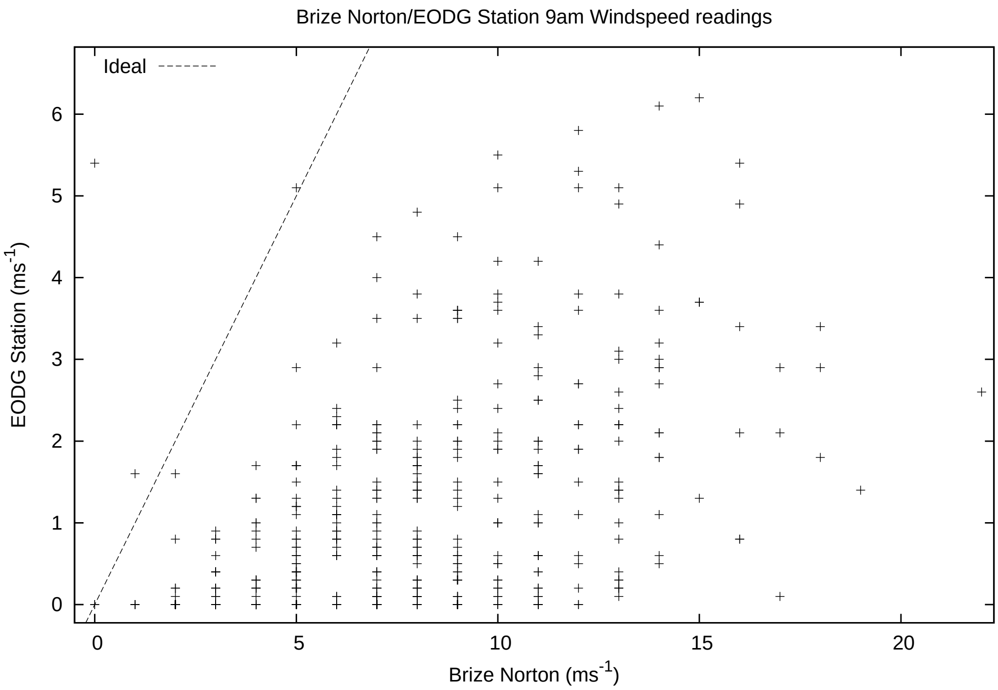

Figure 14: Comparison of 0900 wind speed readings

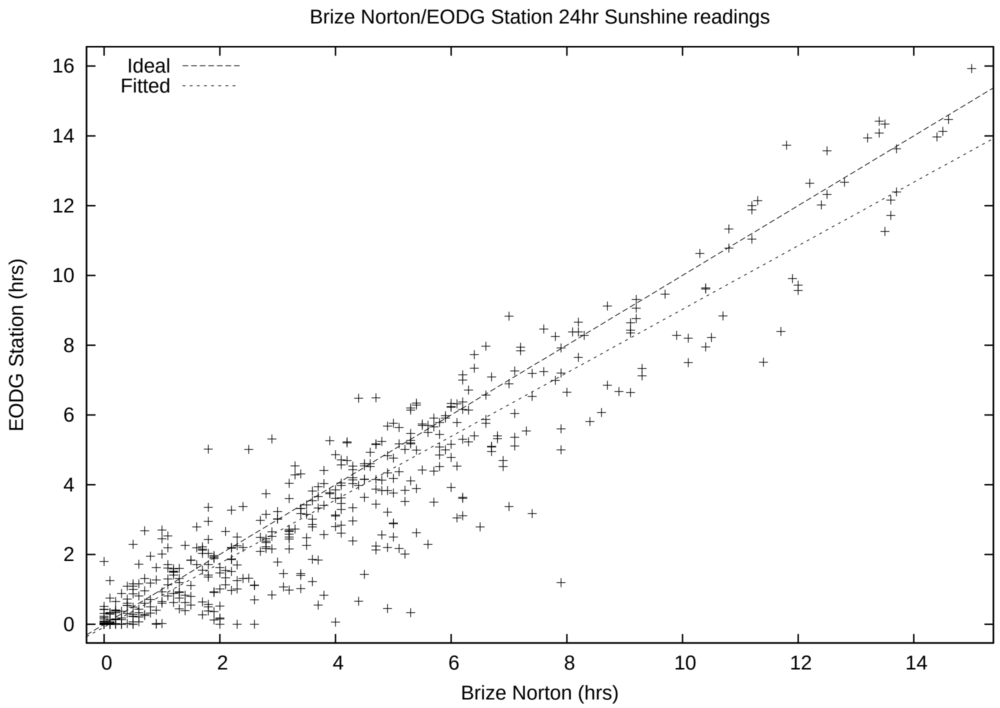

Figure 15: Comparison of 24 hour sunshine readings

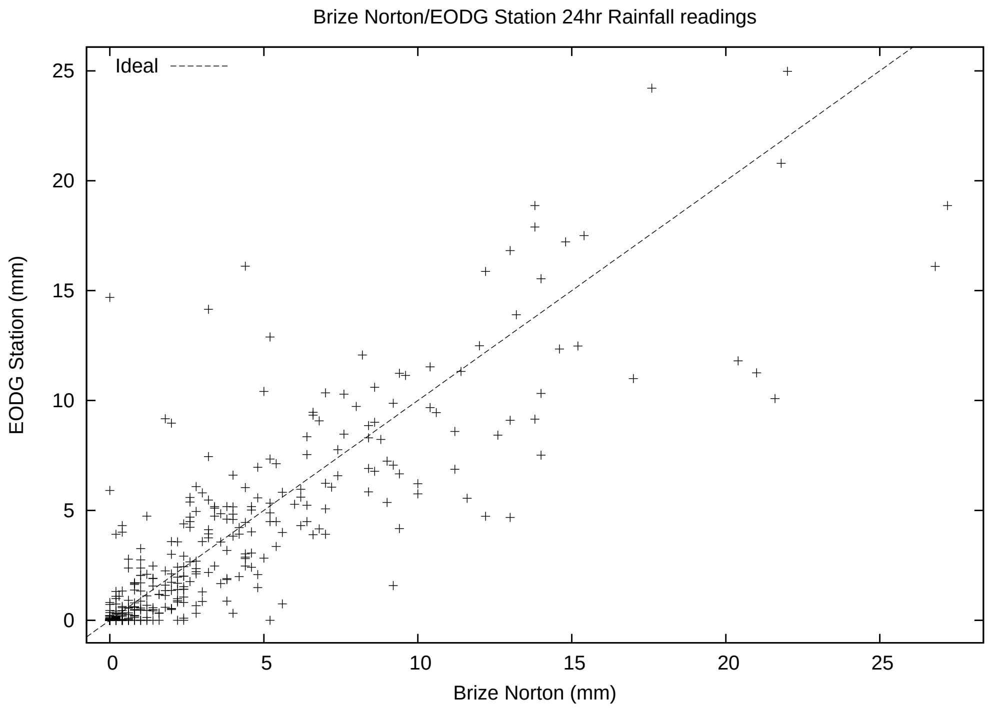

Figure 16: Comparison of 24 hour rainfall readings

5 Direct comparison to collocated sensors

In order to eliminate problems attempting to compare weather data across different exposures and locations we decided to build a cost effective sensor package that could be mounted next to the existing EODG station and provide an additional source to compare to. One method of obtaining data might be to take some portable handheld sensors onto the roof and take periodic readings, but this would be a very time consuming way of gathering not much data, and readings would also not represent all weather conditions, as access to the roof is only safe in good conditions in daylight; so an automated station constantly taking readings is preferable.

Although we could acquire another weather station or set of atmospheric sensors of similar type to the existing station; this would cost too much for the project; alternatively one could integrate off the shelf sensors and and a datalogger separately but this is still too expensive. Therefore by buying the raw sensors and circuit components and a small microcomputer, integrating them into a complete system and writing all of the code, one can obtain a much more accurate and precise system for a smaller amount of money - just with more work. Additionally, by going direct to sensor components from the manufacturer we can get a much better idea of the typical accuracy of each sensor.

5.1 Design of Raspberry Pi based weather sensor package

The system is based on a Raspberry Pi, a single board Linux based computer, which provides a high level programming environment (Python) for ease of development, easy access to the wireless network for remote access by way of a USB wireless dongle, and also access to low level digital inputs for sensors. Another alternative considered was a lower level microcontroller such as an Arduino or PIC based system, which would be slightly easier to interface to the sensors but almost impossible to download data from without going onto the roof, or buying an expensive wireless interface. Using a single board computer enables remote access easily with a cheap USB dongle, as it can simply be connected to the Physics department wireless network as any other computer, and accessed via ssh. The complete system shall henceforth be known as "PiWX".

A block diagram of the system is shown in Figure 17, and a summary of the materials used is shown at the end of this section.

The first sensor is a combined temperature/pressure solid state sensor providing a reasonable degree of accuracy and precision that has been factory calibrated. The second sensor is an electronic humidity sensor. The current EODG station reports hours of sunshine per day but only uses an on/off sensor to do this, and the light level at which it triggers is unknown - the World Meteorological Organisation standard is 120Wm^{-2} but it is unclear if the sensor has been calibrated to this level. Therefore a variable light sensor that outputs a continuous range of values was attached so that the this level could be adjusted. The light sensor outputs in units of lux, but it is possible to convert from lux to Wm^{-2}, which will be done later in the data analysis.

For windspeed and direction, I wanted to use a spare cup anemometer and direction vane from the previous weather station that could be interfaced with the Pi - these sensors are analogue and so 2 channels of analogue to digital conversion are required. However, when the wind sensor was attached, it was discovered that it had been damaged at some stage, and no longer gave reliable readings of either wind speed or direction. Upon consulting with the manufacturers of the anemometer it proved uneconomical to repair, and unfortunately I did not have the budget to buy another one. Therefore it proved impossible to provide an alternative source of data for the wind; this will be discussed further later.

I also decided that there was no need to try to measure the rainfall; as the existing rainfall gauge could easily be directly calibrated by pouring measured amounts of water into it.

Although it appears that very little was spent on sensors compared to the datalogger system; the datalogger system based on the Pi is essentially a fixed cost regardless of the quality of sensors plugged into it. Moving from this grade of sensors to the next sort of price step unfortunately involves several orders of magnitude of cost; however even at this price level the sensors are individually factory calibrated.

I wrote a basic automated datalogger program that works in much the same way as the existing EODG Station, as it averages readings over 10 seconds and writes them to a csv file for each day.

Finally the system was physically packaged and weatherproofed; the temperature, pressure and humidity sensors obviously needed to be in a vented case and the computer needs to be in an attached waterproof case. I reused a waterproof case from a previous project in the department, and bought a small plastic case to attach to the side of it for the sensors, which I drilled multiple holes in so it was

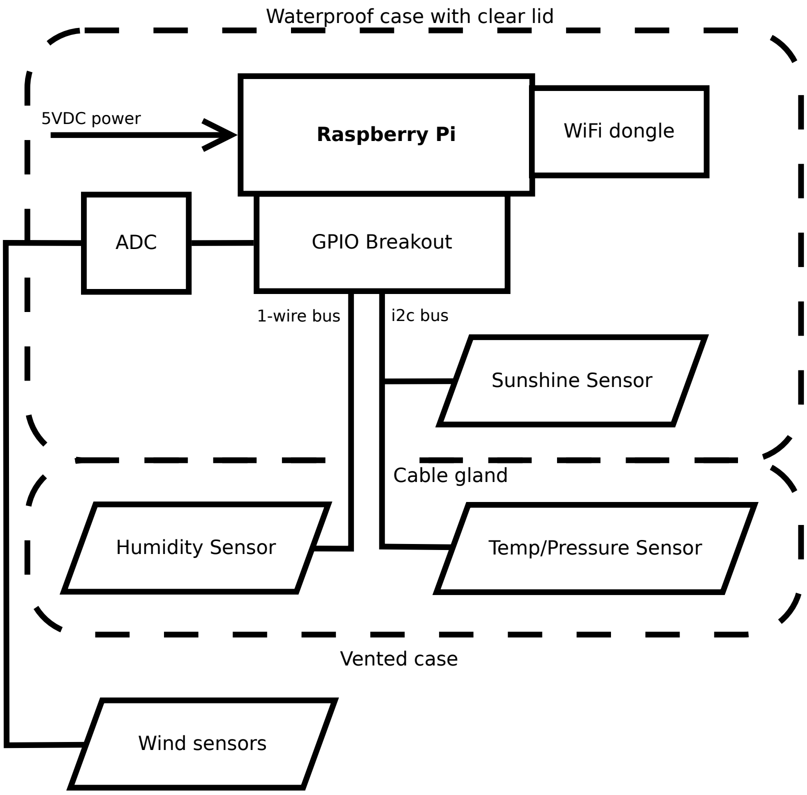

Figure 17: Original System

vented, but the sensors were still protected from direct sunlight and rain.

5.1.1 Power Supply

One of the challenges of putting the device on the roof was getting power to it in a safe way; the Pi requires 5VDC and draws a maximum of 2 amps. The cable run from the nearest indoor socket, through a cable trunk, to the roof, is at least 10 metres. Any reasonably sized cable will suffer from a large voltage drop over that length. Alternatively 24VDC could be supplied at a lower current, then a DC-DC converter placed at the Pi to take it back to 5V. However this is complex and requires using more components.

There is a waterproof mains supply on the roof available on a 16 amp CEEform connector - after consulting with the Central Electronics Group I decided to utilise this by taking the mains supply into a separate weatherproof and waterproof box already approved for this use by the CEG, and then terminating with a normal 13A socket, into which the normal Pi power supply will be plugged. This eliminates any problems with voltage drop, and minimises the amount of mains powered equipment I will have to work with. It also ensures that all mains electricity is physically separated from the low voltage electronics I constructed in a separate enclosure, so the enclosure I constructed did not have to be as rigorously tested.

5.1.2 Safety

When constructing electronics, particularly mains powered equipment, safety is obviously paramount. The most obvious risk during all project stages was electrocution - while most of the project uses safety extra low voltage equipment(Class A as defined in the Oxford Physics Health and Safety manual, maintained by the Electrical Safety Advisory Committee) at 5VDC at 2A or less, which is generally considered touch safe, the Raspberry Pi is powered by a power adaptor which attaches to the mains power supply. This will be an off the shelf unit, however I built the 16A-13A plug adaptor in order to use the power supply on the roof - I have built such adaptors before and CEG were happy for me to build it myself - this and the DC power adaptor were PAT tested in line with Physics department policy.

During the build there was also a small risk of solder burns etc., but I have experience working with electronic test equipment and soldering equipment both as part of the Oxford Physics Practical course and also externally when I have worked as a theatrical technician, feel safe using this equipment and took all standard lab safety precautions.

Additionally, the 16A outlet on the roof is on an individual circuit breaker and all fall protection measures were still in place from when the outlet was used for a LIDAR project. Any roof access was conducted during office hours in good conditions - access to the roof for this project was subject to a separate risk assessment from the start in any case for access to the existing EODG Station.

5.1.3 Materials used

	Object	Supplier
Base	Raspberry Pi B	thepihut.com
	USB WiFi adaptor	thepihut.com
	8Gb SD card	thepihut.com
	Humble Pi GPIO breakout	
	Aux stripboard	thepihut.com
	MCP3008 8 channel ADC	thepihut.com
Sensors	Adafruit DHT22 humidity sensor	thepihut.com
	Adafruit Lux sensor	thepihut.com
	BMP180 temp/pressure sensor	Cool Components
	Wind sensor	reused
Mechanical	IP54 ABS sensor case	Maplin
	IP65 Weatherproof main case	reused
Power	IP67 Waterproof 16A connector	A1 Electrics
	Outdoor socket box	Farnell/Onecall
	Cable glands	Maplin

5.2 Initial Results from PiWX

Once constructed the PiWX system was installed on the roof near the existing sensors. As remote access capability had been built in data could be analysed right away - although the software and hardware all worked very well at first, it was clear after a few days that not all was well with the sensors placed in the vented section (temperature, pressure, relative humidity,) particularly the temperature sensor. Although the sensors had been placed in the shielded box specifically to take them out of direct sunlight, it appeared that sunlight was making the inside of the box heat up, thus affecting the temperature and relative humidity readings. This can be seen in figure 18, as the EODG Station temperature and the PiWX temperature track each other well during the night, but during the day the PiWX temperature quickly rises to improbably high temperatures. Additionally, it was discovered that the light sensor was reading off scale high when in direct sunlight over 40000 lux, hence the parts of the graph where the light level drops suddenly to zero despite being in the middle of the day. It was decided that as greater than 40000 lux represented a level clearly above the WMO standard for sunlight ($120\text{Wm}^{-2} = 11160\text{Lux}$ for sunlight, a conversion factor that will be discussed later) then the times when the light level dropped to during the day would be countable as sun, and this change would be applied during data analysis, so no change was made to the software.

One could of course note that as we are attempting to calibrate the EODG Station sensors from the Pi there might be reason to suggest that perhaps the PiWX sensors were indicating the true value, and it was instead the EODG Station sensor not warming up enough in the day. However, the previous comparisons to MIDAS and the Radcliffe Observatory showed that the EODG temperature sensor was indicating the correct values to within at least a few degrees, so it is clearly the PiWX sensors that are wrong.

It was apparent that the PiWX temperature and relative humidity sensors were so inaccurate as to

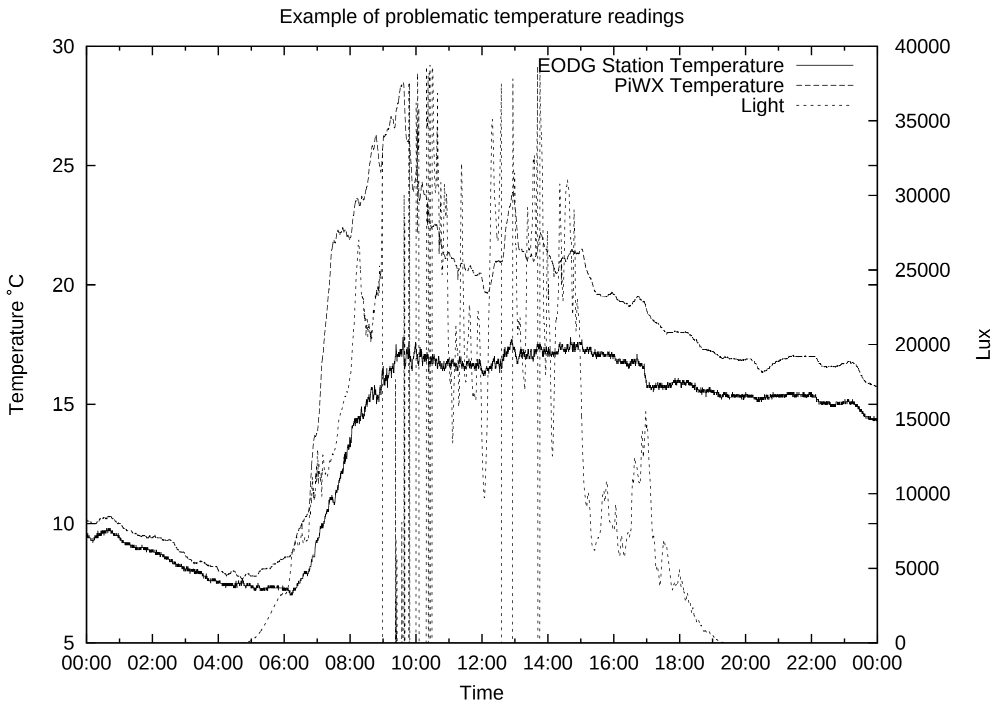

Figure 18: Example of problematic temperature readings

be useless, much less able to calibrate the EODG Station, so the PiWX system was removed from the roof and redesigned slightly.

5.2.1 Design Changes

The main design change was removing the sensors from the vented box attached to the Pi, and putting them into a small Stephenson screen assembly open to the air. This greatly increased air circulation through the sensors, and also reflected solar radiation more effectively, as the screen was made of a light coloured thin plastic rather than the previous thick grey plastic that absorbed and trapped heat. I also took the opportunity to use a long cable I had originally been intending for the wind sensors (which had never been mounted, as they did not work) in order to place the sensors immediately below the existing EODG sensors on the pole. The Pi was then reinstalled and reactivated; after a few days the data was reexamined and the effect of solar radiation appeared to be completely eradicated, with the temperature sensors tracking each other well.

5.3 Results

The PiWX system was left collecting data for an 18 day period. As this was in September no periods of sustained heavy rain, snow or other extreme weather conditions were experienced, but certainly a range of temperature, pressure and humidity values were collected that are representative of most of the typical range. Despite the short timescale of the comparison compared to the MIDAS/Radcliffe Observatory comparisons, which compared across a timescale of years, a similar number of data points was obtained by taking a data point every 10 minutes, rather than only once every day. This also had the advantage of comparing the sensors all times of day, rather than only at 0900GMT, which might have exposed any time-dependent issues (none were found.)

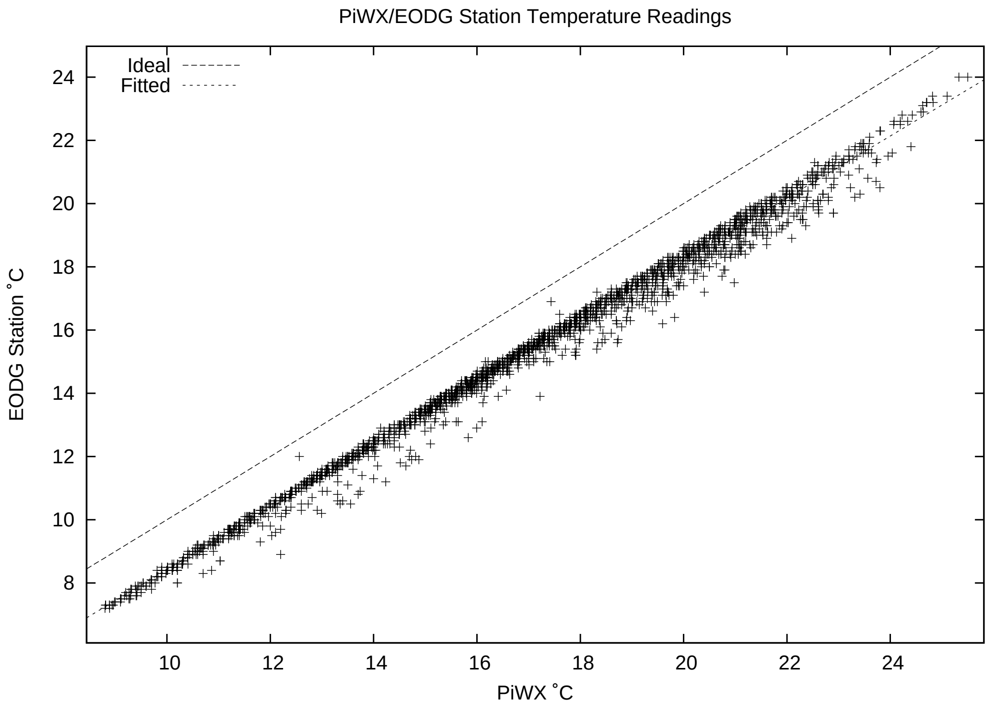

Figure 19: Comparison of temperature readings

5.3.1 Temperature

Figure 19 shows the comparison of the temperature readings of PiWX and the EODG Station at 10 minute intervals over the data collection period. It is obvious that the data is well correlated, and a similar linear fit to those performed for the MIDAS and the Radcliffe Observatory comparisons yields $a = -1.3728 \pm 0.02841$ for the constant term, $b = 0.97965 \pm 0.00179$ for the linear. These are very similar to the previously obtained values adding additional confidence that we have correctly obtained calibration coefficients.

However, as mentioned earlier, we actually know the stated typical accuracy of the PiWX instruments from the manufacturers datasheets, which we did not know for the MIDAS or Radcliffe Observatory comparison, and so this can also be used as input for the fit, making the uncertainty on the fit parameters much more useful. In this case, the manufacturer states that the typical accuracy of the temperature sensor is $\pm 0.5°C$. However, thus far we have plotted the EODG Station on the y axis and the comparison data on the x axis, with the implicit assumption that all error was on the y axis, and the comparison data represented the 'true' values. However, we now actually *know* the error on the x axis. This is mildly annoying because the implementation of curve fitting the plotting and fitting package I am using (Gnuplot) only takes account of errors on the y axis. Thus, to fit using the known uncertainties I inverted the axes and also inverted the function for the fit, but have continued to show the graphs and fit parameters as before for ease of comparison.

5.3.2 Pressure

Figure 20 shows the comparison of the pressure readings of PiWX and the EODG Station. The granularity evident in the EODG shows the much greater precision available on the Pi sensor. The absolute accuracy of the PiWX pressure sensor is given as ± 2.0mb, and so the fit parameters were generated in the same manner as described for the temperature to take account of this. Additionally, both sets of pressure readings were corrected for height, so these values are pressure at MSL (this was necessary as the PiWX pressure sensor was mounted on the pole on the roof, but the EODG Station sensor is inside the building next to the datalogger, giving a few meters height difference, but it is unlikely that the

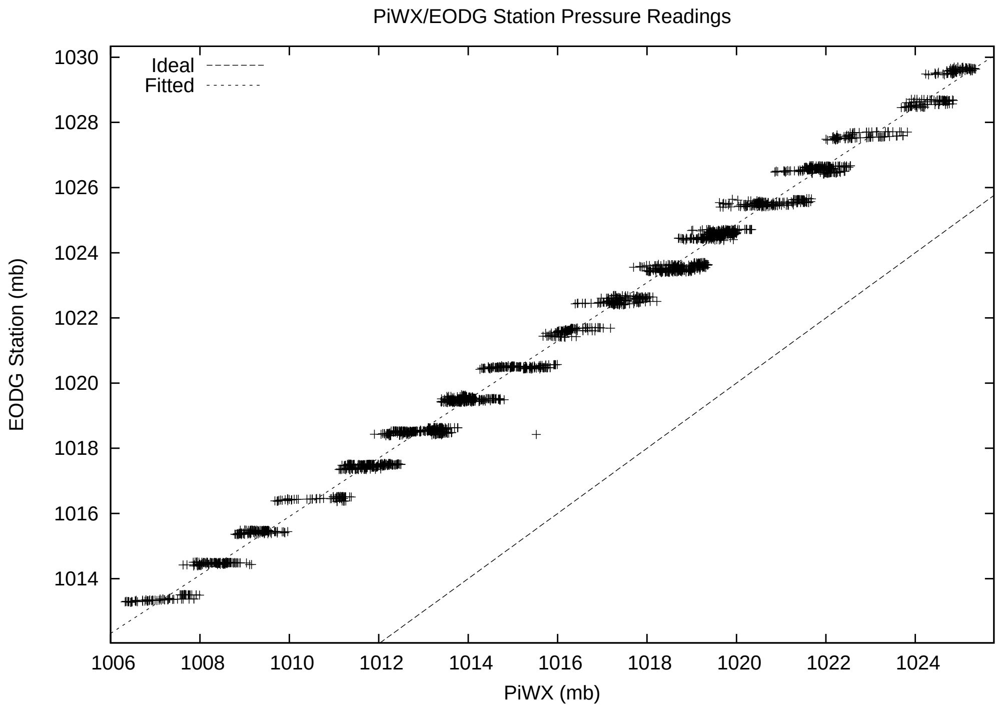

Figure 20: Comparison of pressure readings

different exposures caused any other difference due to pressure only varying over long lengthscales and timescales.)

The linear fit parameters to this data are a $= 109.43 \pm 1.573$ for the constant term, and b $= 0.89750 \pm 0.0017$ for the linear. These are again similar to the previously obtained values from the Radcliffe/MIDAS comparisons; the constant term is a little lower but is still within the previously given fit uncertainties - additionally, this value has a much smaller uncertainty due to the instrumental uncertainty being known.

5.3.3 Relative Humidity

The comparison of relative humidity readings is shown in figure 21. This graph shows very good correlation when compared to the MIDAS/Radcliffe comparisons, although it does appear at the top end of the range that the PiWX sensor is hitting the top limit of 100% too early - this could be a fault of the physical sensor package allowing humidity to build up inside the Stephenson screen; but as that is exactly what the screen is intended to prevent this seems unlikely - the MIDAS comparison did appear to show that the EODG station was undereading slightly at the top end of the range but the fact that the Pi sensor, for whatever reason, is sticking at 100% (usually in the early morning) is odd.

The linear fit parameters to this data are a $= -0.61856 \pm 0.6121$ for the constant term, and b $= 0.93512 \pm 0.0020$ for the linear.

5.3.4 Sunshine

As mentioned previously, the light sensor chosen output in lux, a unit which is a measure of intensity as perceived by the human eye. We need to convert this into a value in Wm^{-2}, power per unit area, or the irradiance. We can do this by integrating across all wavelengths, multiplying the spectral density function of the source, B_λ in fractional contribution per wavelength(in this case, sunlight, which can be

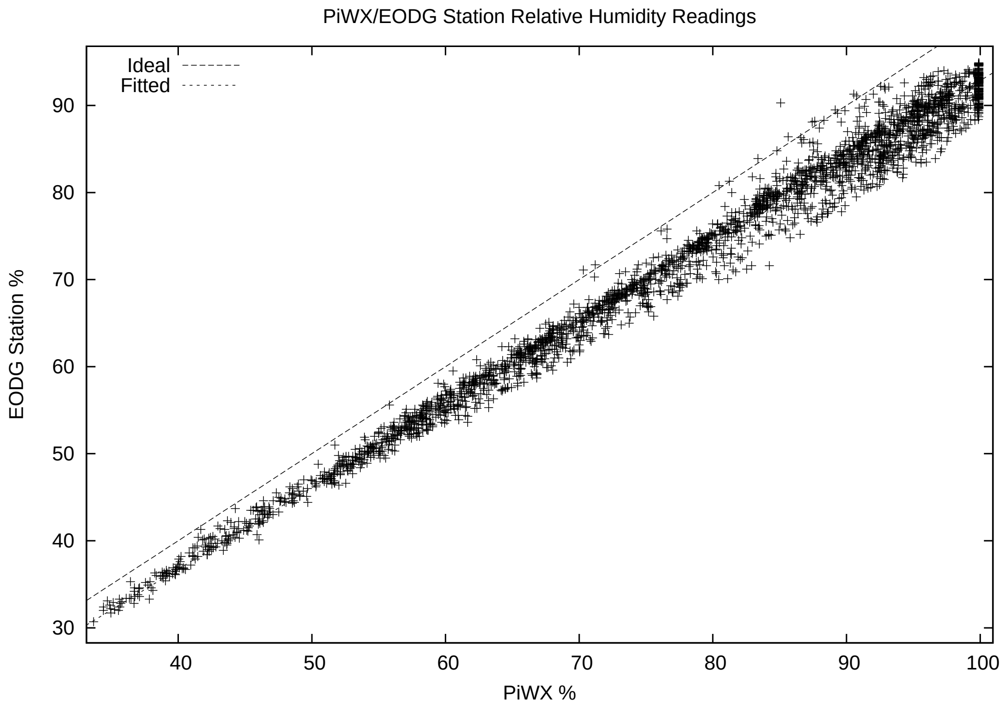

Figure 21: Comparison of relative humidity readings

represented as a black body at 5800K) with the standard CIE photopic sensitivity curve of the eye $\bar{y}(\lambda)$:

$$\epsilon = 683 \int_0^\infty \bar{y}(\lambda) B_\lambda \mathrm{d}\lambda$$

This generates a conversion factor, or luminous efficiency, in units of $\mathrm{LuxW^{-1}m^2}$. For a blackbody at the temperature of the surface of the sun the integral evaluates to 93, so all one has to do to get the irradiance is to divide the lux values by 93. As mentioned previously when the light level was greater than 40000 Lux the Pi light sensor tripped off scale high and output a value of 0, so as this was well over the WMO definition of sunlight any readings of 0 during the day were counted as sunlight. At this stage another problem became apparent - the readings of sunlight from the EODG station and the PiWX sensor seemed completely uncorrelated. As an example the readings for a single day are shown in figure 22, and most days take the same form: one can see that there are periods where the EODG station recorded sunlight, despite the PiWX sensor recording less than $120\mathrm{Wm^{-2}}$, periods where EODG sensor has both recorded and not recorded sunlight despite the level from PiWX being consistently over $120\mathrm{Wm^{-2}}$.

It is unknown if this is a problem with the Pi sensor, the EODG sensor or more possibly both; it is possible that in very sunny conditions(but at a different level to the Pi!) that the EODG sensor also trips off scale high and actually reads 0. Both the previous comparisons to Brize Norton and the Radcliffe Observatory show that the sunlight sensor tends to under read the amount of hours of sunshine in the day, so this could be a possible explanation for that behavior. Without further investigation of the way the EODG sensor works it is difficult to conclude anything further from this data. The hope was to be able to use the Pi's variable light sensor to find the level that the EODG sensor tripped on at and adjust this if needed, but with the data available identifying a consistent level at which it does this has proved impossible.

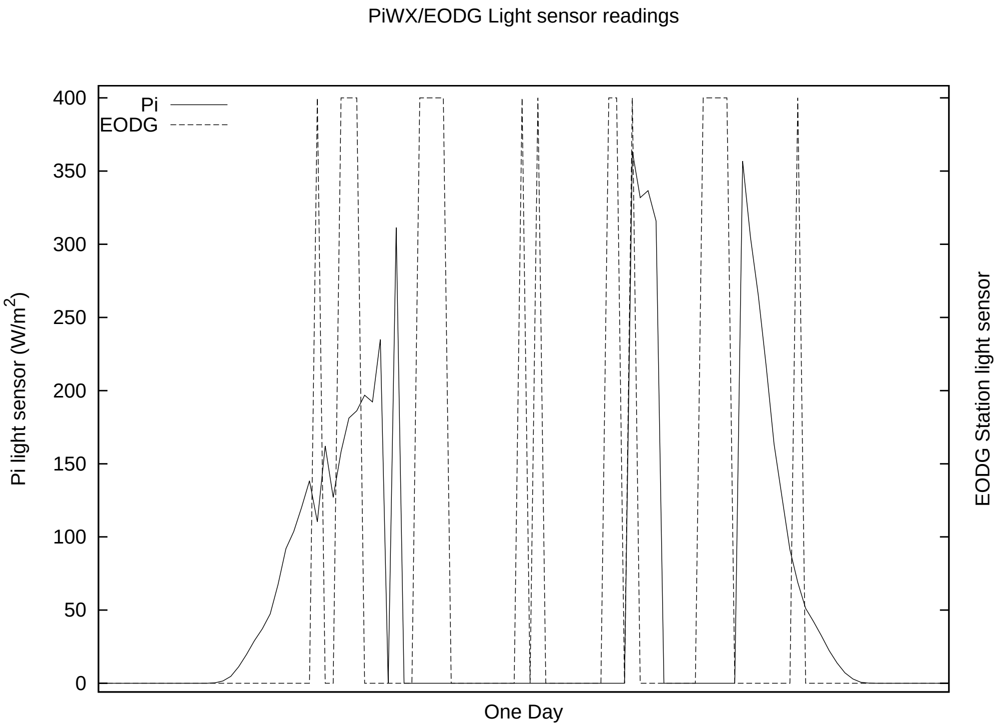

Figure 22: Example of problematic light sensor readings

6 Direct calibration of rainfall and wind direction

6.1 Rainfall sensor

No attempt was made to provide a rainfall sensor for PiWX due to the ease of providing a direct check on the calibration of the existing EODG sensor by pouring measured amounts of water into it. Rainfall is typically measured in millimetres, which is of course the level of 1 litre of water distributed over an area of 1 m^2. The diameter of the inlet of the rainfall sensor is 8cm, so the area it covers it $0.0050265m^2$. This means that in order to record 1mm of rain, 5.03ml of liquid should have entered the sensor. The smallest graduation the sensor records is 0.01mm, or 0.05ml. Unfortunately finding a pipette or syringe that can measure accurately down to a twentieth of a millilitre in a physics lab proved difficult, and so larger amounts were used, with an accuracy of 0.5ml, or about 0.1mm once converted to millimetres. Various amounts of water were dropped into the sensor from a syringe; at first the whole amount was poured in at once but the sensor was not designed to cope with this much water at once (as, obviously, a rainfall rate of 1-2mm in about half a second indicates that the Atmospheric Physics building will be completely submerged within a few hours, and exact meteorological readings are unlikely to be of concern in this scenario,) so the water had to be added a droplet at a time, as in real rain. This was very time consuming and so not as much data was gathered as I would have liked, but the results are shown in figure 23. The errors being on the x axis data were once again dealt with using the same procedure as before to use them in the linear fit. The fit parameters are a $= 0.0801995 \pm 0.04011$ for the constant term, and b $= 0.896787 \pm 0.04498$ for the linear term.

It does appear from looking at the data that the majority of the data points are close to the ideal 1:1 line, but in particular the last set of data points wanders downwards, which has pulled the fit gradient down. It is unknown why the last set of repeated readings exhibited this phenomena.

6.2 Wind direction

As I had been unable to get the spare wind sensor to work, I decided to attempt calibration of the existing wind direction sensor by temporarily rotating the pole down into reach, rotating the wind direction vane

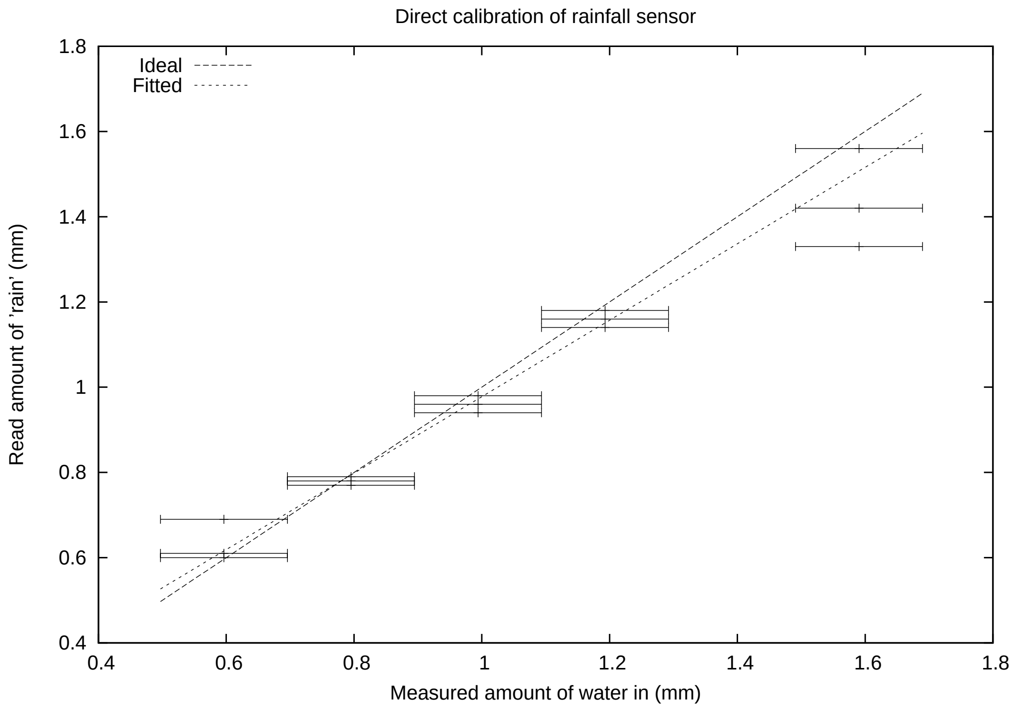

Figure 23: Direct calibration of rainfall sensor

to point at each direction and then comparing those values against the values the datalogger recorded. To do this, I printed a paper protractor, stuck it onto the top of the wind direction vane, and attached a pointer to the structure of the wind sensor pointing at the protractor so that I could reliably set directions. I decided to measure every 30 degrees, as from the small size of the divisions on the top of the wind direction vane and the somewhat improvised nature of the setup, it was probably not accurate to more than about 5 degrees.

As can be seen in figure 24 the read values do seem to track the value input by rotating the vane quite well a linear fit to the data produces the parameters a $= 6.14853 \pm 1.405$ for the constant term, and b $= 1.03606 \pm 0.007051$ for the linear term. Although there is a small non-ideal linear component, most of the difference seems to be a constant offset, which for a circular variable like angle would be an unusual error - it seems more likely that the paper protractor I was using to measure the angles was stuck onto the vane at a slight angle from true - it was only stuck on with blu-tack, and there was no other way other than lining it up by eye.

7 Other work completed as part of the project

7.1 Online interface to data

As well as investigating the calibration of the EODG Weather Station's sensors and building the Pi weather station, some other work was undertaken as part of the project, mainly related to the website which displays the data from the weather station. Although live readouts and graphs of today/yesterday are immediately available, long term data is only provided in the raw data format output from the datalogger, which takes the form of a csv file for each day. This format, while helpful for detailed investigation, is not user-friendly to the casual browser or for easily viewing long term trends in the data. Therefore, I decided to build a visual interface to the data in the form of user-controllable graphs that could view any time period, and see maximums, minimums, average trends etc.

Unfortunately, due to the large size of the dataset (440MB at time of writing) attempting to render graphs dynamically in a web browser quickly proved to be an exercise in futility, as Javascript was not

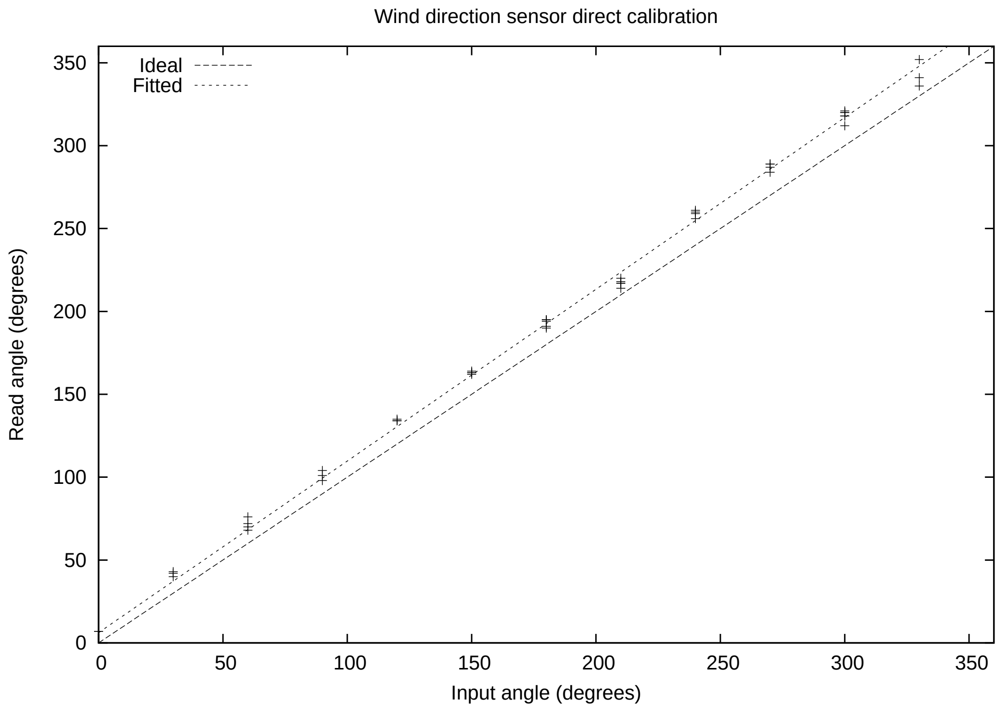

Figure 24: Direct calibration of wind direction vane

really intended to handle that much data in a graphical format. Instead I went with a slightly less flexible but much easier to build system; the graphs are rendered offline into images for each time period and are loaded as needed. This means the user can select to view either a day, week, month or year, or indeed the complete dataset from the commissioning of the weather station onwards. What the user cannot do using a static system is select to view a time period starting at an arbitrary point - they must select a day as runs from midnight-midnight, or a calendar month rather than a month's worth of time starting midway through a calendar month. In practice this restriction will likely not cause any problems. Additionally, after zooming out further than a week only the daily maxima, minima and average values are plotted for each parameter, as viewing the detailed time series would give a very cluttered graph at long time scales.

This page is viewable on the EODG website[1] and is automatically updated with the latest data overnight.

7.2 Air pollution data

Oxford City Council make available data on measurements of air pollution taken by automated monitoring sites around and outside Oxford as part of the 'Oxford Airwatch' project, which determines by how much Oxford has missed air pollution reduction targets each year. Although the yearly report produced by the Council compares pollution levels to national and local targets, it does not investigate any possible connection with the local weather. The pollution data is given on an hourly basis, with various pollutants measured such as Nitric Oxide, Nitrogen Dioxide, volatile and non-volatile PM10/PM2.5, and Sulphur Dioxide. It is easy to scan the EODG Station's historical dataset and to produce a similar set of hourly readings of the weather, and thus compare the two to investigate any possible correlations between, for example, wind direction and particulate matter levels, which might help to identify what the sources of pollution actually are.

Unfortunately despite much searching very little data supporting links between pollution levels and the weather was found in this combined dataset. For example, the Sulphur Dioxide sensor at Harwell,

[1] http://eodg.atm.ox.ac.uk/eodg/weather/explore

outside the city, registered on average several times higher readings when the wind was blowing from about 60°than anywhere else, but is clearly due to Didcot Power Station, which is on that bearing from Harwell. There were some noticeable trends in relation to the local Ozone measurements, which tended to increase at high temperatures, and decrease at higher relative humidities, but these are known general climate effects and are not specific to Oxford. No other definitive trends or correlations could be spotted when compared to the weather, so no further work was conducted on this data.

8 Conclusions

Over the calibration project, several sets of comparisons have been conducted to several different data sources to attempt to confirm the integrity and accuracy of the sensors attached to the EODG Weather Station. A summary table of the various calibration coefficients generated for each sensor against the comparison data source is presented below:

EODG sensor		Radcliffe Observatory	MIDAS/Brize Norton	PiWX or Direct Calibration
Temperature	a	-1.43614 ± 0.05907	-0.654272 ± 0.06487	-1.3728 ± 0.02841
	b	0.971807 ± 0.005137	0.967576 ± 0.005988	0.97965 ± 0.00179
Pressure	a	123.463 ± 14.52	126.462 ± 13.9	109.43 ± 1.573
	b	0.8818 ± 0.01433	0.871405 ± 0.01373	0.89750 ± 0.0017
Rel. Humidity	a		2.34902 ± 1.36	-0.61856 ± 0.6121
	b		0.960655 ± 0.01652	0.93512 ± 0.0020
Sunlight	a	0.110767 ± 0.09836	-0.0807917 ± 0.06483	
	b	0.741107 ± 0.01726	0.910935 ± 0.01308	
Rainfall	a	0.0762897 ± 0.0444		0.0801995 ± 0.04011
	b	1.00209 ± 0.008999		0.896787 ± 0.04498

Not much reliable calibration data has been obtained on the wind sensor due to the difficulties outlined; the direct calibration does imply that the wind direction sensor is working correctly and is calibrated correctly to within a few degrees; however the wind speed sensor is still of essentially unknown quality.

Calibration of the other sensors has been more successful however. Taking first the temperature sensor; the coefficients obtained by comparing to PiWX and to the Radcliffe were virtually the same - the Brize Norton coefficients were a little different but that is to be expected given the much greater distance to the airfield. As the Pi was situated directly next to the EODG Station I decided to take it's coefficients for this sensor.

For pressure, all three data sources found very similar coefficients, but the Pi's had smaller uncertainties due to a high precision sensor with a known accuracy being used, and so I decided again to recommend the Pi coefficients. Relative Humidity was a little tricker, as the Radcliffe Observatory was itself dubious at the high end of the range. As the Pi data was here better correlated than the Brize data, I decided to recommend the Pi's coefficients for this sensor also.

Sunlight is also difficult, as attempting to calibrate the sensor on/off level using the PiWX sensor failed as described. The two comparisons of number of hours of sunlight recorded per day to the Radcliffe Observatory and Brize Norton resulted in quite different calibration coefficients being found, but both indicate some degree of under-reading. With this limited data and uncertain performance of the sensor it is difficult to recommend a specific calibration function but the data should be treated with caution.

The rainfall data on the other hand is a little easier; no fit was attempted to the Brize Norton comparison as the data was too clustered and uncorrelated to get a reliable fit, but the comparison to the Radcliffe Observatory indicates an almost ideal calibration. The direct measurements performed upon the rainfall sensor have considerably less data points than is optimal; the long time required to take each data point reduced the amount of data taken. Looking at the fit it appears that two low data points right at the end are pulling the fit away from ideal, which could be an error. With this in mind I will tentatively recommend that no calibration is needed for the rainfall sensor.

8.1 Recommended Calibrations

The functions given below are formulated so as to calibrate a recorded value, x, into a calibrated value.

EODG sensor	Recommended function	Comments
Temperature	$= \frac{x+1.3728}{0.97965}$	
Pressure	$= \frac{x-109.43}{0.8975}$	
Rel. Humidity	$= \frac{x+0.61856}{0.93512}$	
Sunlight		Treat data with caution, probably under reads true value
Rainfall	$= x$	None required, but more direct measurements required to confirm
Wind speed		Readings very influenced by exposure, comparisons to nearby sources show a significant under-reading
Wind direction	$= x$	Direct measurements indicate accuracy seems to be within 5 degrees

8.2 Thanks

Thanks must be extended to the Physics Department for funding this summer project, the Earth Observation Data Group for being my gracious hosts for the summer, and in particular Dr. Adam Povey for supervising the project.

9 Further Work

Although calibration coefficients have been confidently established for several of the key EODG Station sensors, others proved difficult to pin down; in particular the wind sensor has had no direct calibration at all due to the difficulty of cheaply sourcing an accurate, working wind speed sensor to attach to the Pi. Additionally, the comparisons of wind speed to the Radcliffe Observatory and to Brize Norton through MIDAS suggest that the EODG sensor consistently under reads the windspeed. It is currently unknown how much of this is due to the sensor being miscalibrated and how much of this may be due to the exposure of the sensor being non-ideal. It is clear that despite being on top of a pole on top of a building the surrounding environment still influences the wind speed considerably - the nearest building to EODG is a few meters higher and the wind is very rarely read as coming from that direction, indicating it acts as a block. Anecdotally standing next to the EODG station and observing the anemometer on top of the much higher Engineering building that the Radcliffe Observatory wind data is taken from, the EODG anemometer is often not rotating when the higher one is. Further work should therefore include studying the wind sensor in particular.

Additionally, the problems presented earlier with the light sensor mean that further work should also focus on determining how the light sensor works, and if it does indeed trip off scale high leading to the anomalous behavior observed. It would be best that the on/off sensor was replaced with a sensor that could read out a continuously variable level, but it is uncertain if this could be integrated into the existing EODG datalogger.

Some more data collection should also be performed on the rainfall sensor; each data point is very time consuming due to the requirement to drip the water in very slowly from a syringe, but this would be a relatively easy way to improve the confidence in the sensor.