Chapter 2

Methodology for Calculating Humidity

SUMMARY

The relationship between temperature and humidity and between humidity variables themselves is non-linear. There are various algorithms for converting between humidity variables, with varying degrees of sophistication and precision. All humidity data used in this project originate from dewpoint temperature in the data base (Chapters 3 and 4) and necessarily have to be converted to get vapour pressure, specific humidity and relative humidity. This Chapter describes the formulae chosen for converting to each of the key variables and the reasoning behind it.

2.1 VAPOUR PRESSURE

There are numerous formulae for calculating e from T_{dw} (and e_s from T) with varying degrees of complexity and accuracy tailored to specific end use. The vast majority are approximations with unavoidable inaccuracies, especially at temperature extremes. The Goff & Gratch (1946) formula has been highly reputed in terms of accuracy and is still in common use (Alduchov & Eskridge, 1996), but they are relatively computationally complex (Murray, 1967). The much simpler Magnus Equation (Magnus, 1844) is one of the most convenient forms:

$$e = c \exp\left(\frac{a\,T_{dw}}{b + T_{dw}}\right) \qquad \text{Eq. 2.1a}$$

where a, b, and c are constants and T_{dw} is in ^{o}C. There are numerous values published for these constants (Wexler, 1976 and 1977; Buck, 1981; Abbot & Tabony, 1985;

Alduchov & Eskridge, 1996). There are similar but slightly more complex formulae with added constants and an enhancement factor which attempts to account for using the equation with moist air rather than pure water (Buck, 1981; Alduchov & Eskridge, 1996).

Two Magnus type and two non-magnus type (more complex) equations have been considered for potential use here and are shown in Table 2.2 along with applicable T ranges and maximum error estimates. The latter two include a fourth constant and an enhancement factor. In all four, substituting T_{dw} with T gives e_s. Equations for e and e_s (the basis of conversions to RH and q) differ depending on whether T_w is above (wet-bulb) or below (ice-bulb) 0 ^oC. This is because e decreases more rapidly over an ice surface than over water (Barry & Chorley, 1998). Equation 2.1b is chosen for its suitable range and low maximum error. Equation 2.1c is chosen as it is frequently referenced and used by Mitchell *et al*. (2004). Equations 2.1d and 2.1e require a surface pressure (P) value and differ only in their enhancement factors. They are chosen for their greater accuracy at lower temperatures and as equations noted to satisfy accuracy thresholds by Alduchov & Eskridge (1996).

As equation 2.1e is reputed to be of high accuracy (Buck, 1981; Alduchov & Eskridge, 1996), all other equations (wet- and ice-bulb) are compared to it over a range of T_{dw} values and two different atmospheric P levels of 750 hPa and 1040 hPa (Fig. 2.1). The differences between vapour pressure calculated from the high accuracy equation (2.1e) and from all other equations (Eqs. 2.1b, c and d), although always less than 1 %, are consistently larger and negative (underestimated) for equations 2.1b and 2.1c. Equation 2.1d shows much better agreement with 2.1e (within ± 0.1 %). For all equations, differences in e relative to equation 2.1e, are larger at temperature extremes. Considering the very close agreement between equations 2.1e and 2.1d, the latter will be used in this project because it is slightly simpler to implement than equation 2.1e.

2.1.1 Choosing a Source of Surface Pressure Data

As P is included in the chosen conversion equation for e it is important to assess the magnitude of any potential error introduced from using potentially inaccurate (or fixed) P values. Using equation 2.1d, the percentage change in e from changing P alone is consistent over all T_{dw} values although this value differs slightly for T_{dw} values above and below zero because of the two different enhancement factors in use. The difference between *e* at a standard *P* of 1013 hPa and *e* calculated at a series of pressures from 750 to 1050 hPa (Fig. 2.2) ranges from -0.11 % to +0.02 % respectively. These differences are smaller when T_{dw} is above zero (the wet-bulb enhancement factor is used). This works out as a difference from *e* at 1013 hPa of 0.0004 and 0.0003 % per one hPa increase in surface *P* for T_{dw} values below and above zero respectively.

Surface *P* data can be obtained: from the simultaneously observed data where available; from an existing global sea level pressure dataset such as HadSLP2 (Allan & Ansell, 2006) with conversions for station elevation over land; or from the standard *P* with conversions for station elevation over land. The following approximation is used for conversion to station elevation:

$$P = 1013 - \left(\frac{Z_s}{10}\right) \qquad \text{Eq. 2.2}$$

where Z_s is the station elevation in meters. The first case is problematic because pressure is not always simultaneously reported thereby reducing the amount of data available. Additionally, by introducing more variables there are more issues of data quality that need to be addressed. The second case provides a useful option. However, matching each observation with the appropriate grid-box month from HadSLP2 is computationally demanding. Furthermore, each dataset has its own issues of data quality and structural uncertainty and so could potentially introduce added complexity and uncertainty to the data. For example, HadSLP2 overestimates *P* over the southern Atlantic and Indian Ocean mid-latitude regions and requires improvements over the Himalayas (Allan & Ansell, 2006). The third case provides a simple method of obtaining *P* values. However, this is only a viable method should the following assumptions hold true: that climatologically, surface *P* is largely consistent with the concept of a standard atmospheric *P* and conversion to station level where necessary; that surface *P* remains largely constant over time (Trenberth & Smith, 2004); and that discrepancies in the above assumptions and small variations in surface *P* do not substantially affect calculated humidity at the monthly mean anomaly resolution.

To investigate these assumptions, the climatological annual mean sea level pressure (over the period 1974 to 2003 as for HadCRUH) and trends at the 5 ^{o} by 5 ^{o} grid-box level (over the period 1973 to 2003 as for HadCRUH) are created from HadSLP2.

Climatological P (Fig. 2.3) is largely between 1008 and 1020 hPa with small patches characteristic of the sub-tropics reaching 1024 hPa. Thus the approximation of a standard surface P over the oceans and a station level P calculated using elevation over the land gives a potential error of -5 to 11 hPa. Trends in surface P are very small (Fig. 2.4) and for the vast majority of the globe excluding the Southern Ocean south of 45 ^{o}S are less than ± 0.3 hPa 10 yr^{-1}.

Using the percentage change in e (as above) for one hPa increase in P (0.0003 to 0.0004 % hPa^{-1}), the magnitude of potential error in e if calculated using a P value of 1013 hPa as opposed to 1008 or 1024 hPa, can be estimated as 0.002 % lower to 0.004 % higher for actual values and ± 0.00012 % 10 yr^{-1} for trends over the period of 1973 to 2003. While data quality is a major concern in this thesis, the focus is to provide a first version dataset and so at some stages small compromises of data quality will have to be made for computational efficiency and time. For e, any such issues of data quality from the use of standard P and station level converted P are so small that they are considered negligible.

2.2 WET-BULB TEMPERATURE

The T_w value is necessary for calculating e and e_s but only in terms of its sign to decide whether to use the wet-bulb or ice-bulb formula (section 2.1). For the marine data, some observations have T_w as well as T_{dw}. Recorded T_w is used unless it is greater than T (which would imply RH > 100 %). For all other observations it must be calculated. Accurately calculating T_w from T_{dw} and vice versa was traditionally done using psychrometric tables. This is obviously not possible to automate as is necessary for this project. However, because only the sign is of importance here, and not quantitative accuracy, approximate calculations are sufficient. Two calculations have been found and considered here. The first is a very simple approximation of T_w (Mitchell *et al.*, 2004) (Eq. 2.3) and the second is a more complex equation (Jensen *et al.*, (1990) In: http://www.faqs.org/faqs/ meteorology/_temp-dewpoint/) (Eq. 2.4).

$$T_w = (0.6\, T_{dw}) + (0.4\, T) \qquad \text{Eq. 2.3}$$

$$T_w = \left(\frac{(a\,T) + (b\,T_{dw})}{a + b} \right) \qquad \text{Eq. 2.4}$$

where
$$a = 0.000066\,P$$
$$b = \left(\frac{409.8\,e}{(T_{dw} + 237.3)^2} \right)$$

Equation 2.4 requires e and P in hPa. Regarding e, this creates a circular problem where T_w is required to calculate e (with respect to water or ice) but first e is required to calculate T_w. As only the sign of the T_w value is of importance, to solve this problem, a provisional e is first calculated with respect to water assuming a positive T_w. The presence of both T_w and T_{dw} in the marine data, assumed to be of reasonable accuracy in this case, provides the opportunity to test these equations for a range of T_{dw} values and observed versus standard P.

As only the sign of the T_w value is important, the large differences (as a percentage) between calculated and ICOADS reported T_w (Fig. 2.5), especially at low temperatures, are not a problem. Equation 2.4, using both standard P and ICOADS reported P, gives T_w values of the same sign as that of the ICOADS reported T_w without fail, whereas equation 2.3 can be problematic at low positive temperatures. Consequently, equation 2.4 will be used in this project. The variation in P, for the marine data at least, makes no discernible difference to the sign of the calculated value. For the marine data, standard P will be used. For the land data, following the convention used in calculating e (section 2.1), P will be calculated from the simultaneously reported station elevation using equation 2.2. Calculations of e will be made with respect to ice when $T_w \leq 0\,^{o}$C and with respect to water when $T_w > 0\,^{o}$C. Apparently, most *RH* sensors measure *RH* with respect to water even below 0 ^{o}C (WMO, 1996). However, over land it is not possible to know which observations come from which instrument type. While information for marine data does exist, obtaining this to apply error corrections to all marine data is deemed beyond the scope of a first version dataset and this project but highly desirable for future versions. This is discussed further in section 2.6.

2.3 SPECIFIC HUMIDITY

The equation for calculating q is in common use (Peixoto & Oort, 1996; Ross & Elliott, 1996):

$$q = 1000\left(\frac{\varepsilon\, e}{P - ((1-\varepsilon)\, e)}\right) \qquad \text{Eq. 2.5}$$

where ε is 0.622 and e and P are in hPa. There are minor variations regarding complexity of the divisors. In this case, this equation is easily computable and so chosen without the need for testing. It calculates q in g kg^{-1} using the calculated e values and standard (marine) or calculated (land) P values (section 2.1).

Conversions to q are more sensitive to P than conversions to e. For the range of climatological annual mean P shown in Figure 2.3 (1008 to 1024 hPa), the mean (for e values of 0 to 40 hPa) percentage differences between q calculated at a standard P (1013 hPa) and at pressures of 1008 and 1024 hPa are +0.50 % and -1.08 % respectively. These mean percentage differences vary very little with e at only ~0.002 % hpa^{-1}, therefore justifying the use of the mean difference which for an increase in surface P of one hPa becomes a ~0.1 % decrease in q. In terms of trends in surface P (± 0.3 hPa $10yr^{-1}$), this becomes ± ~0.03 % $10yr^{-1}$ in q (increasing P results in decreasing q). For an approximate global mean q of 8 g kg^{-1} (from a global mean T of 14 ^{o}C (Jones *et al.*, 1999) assuming 70 % *RH*) a global trend of 0.06 g kg^{-1} $10yr^{-1}$ (Dai, 2006) equates to a 0.75 % $10yr^{-1}$ increase in q which is considerably larger than the potential errors in q due to using a constant P.

Although larger than for e, these errors are still small. Hourly surface P may vary quite considerably from the mean. However, this noise is greatly reduced by using monthly mean anomalies averaged over 5 ^{o} by 5 ^{o} grid-boxes. For HadCRUH timeseries and climatologies there may be a slight bias in regions characterised by above or below average surface P (e.g. the Azores High). However, it is the recent changes in surface humidity that are of interest for this thesis and as surface P changes very little over the period of record this is not considered to be a problem. Conclusively, the use of a standard P over the oceans and station level P converted from standard P using station elevation is justified here. For later versions of HadCRUH, where quality assessments of other (non-humidity) variables are possible, it may be desirable to use simultaneously reported *P*.

As a further note, *q* is strictly a proportional measure of humidity, relating the mass of water vapour in the atmosphere to the total mass of the moist atmosphere (section 1.4). Although it is theoretically possible for *q* to change due to changes in *P* alone (see equation 2.5), in practice *q* in HadCRUH is calculated using a constant *P* value over time and so any change in *q* can only result from changes in *e*. Hence, in this thesis it is sometimes referred to as a measure of 'actual' humidity where increasing *q* directly implies increasing absolute atmospheric moisture content as opposed to the relative measure of *RH*.

2.4 RELATIVE HUMIDITY

Calculating *RH* is straight forward, using *e* and e_s values:

$$RH = 100\left(\frac{e}{e_s}\right) \qquad \text{Eq. 2.6}$$

where e_s is calculated simply by substituting T_{dw} with *T* in the equation for *e* (Eq. 2.1d). Notably, this is the only humidity variable used that incorporates *T* in a quantitative way. Furthermore, as a potentially directly measured variable in the first instance, it should be noted that calculated *RH* values are likely to differ fractionally from their measured origins due to conversion inaccuracies on the route from *RH* to T_{dw}, *e* and finally *RH* again.

2.5 DERIVING HUMIDITY VARIABLES AT DIFFERENT RESOLUTIONS

Conversion to and between humidity variables at lower than point source resolution (e.g. using daily or monthly mean as opposed to hourly measurements) introduces error (New *et al.*, 2000; McCarthy & Willett, 2006). This error was found to affect decadal trends but become less important as timescales lengthened. For example, trends in e_s when derived from point source hourly *SST* as opposed to monthly were 0.24 % $10yr^{-1}$

more positive over the period 1987 to 2004. This became only 0.06 % $10yr^{-1}$ when calculated over the period 1973 to 2004. To investigate this further, a sample year (1980) of hourly data from station 037720 (Heathrow, UK) is converted to monthly mean *e*, *q* and *RH* from hourly *e*, *q* and *RH* and monthly mean *T* and T_{dw}. From this simple analysis, conversions from monthly as opposed to point source *T* and T_{dw} give consistently lower humidity values. Averaged over the annual cycle, the mean errors (as a percentage) for conversions at monthly mean compared to hourly resolution are 1.9, 2.1 and 1.8 % for *e*, *q* and *RH* respectively. Considering the large potential error in absolute values and likely errors in trends, all variable conversions for HadCRUH are made at the original observation level.

2.6 THE SOURCE VARIABLES AND POTENTIAL ERRORS

Although both the land and marine data bases report humidity as T_{dw} this has most likely been originally observed as T_w or *RH* and derived accordingly. There is no worldwide agreed convention for converting between these variables. Algorithms, hygrometric tables and psychrometer coefficients can vary within political boundaries and even at a single station over time with very little data available over land regarding what was measured, how it was converted and when any changes might have occurred.

To investigate the possible error this might introduce a sample year (1980) of hourly *T* and T_{dw} observations are taken from station 037720 (Heathrow, UK) and converted to hourly *e*, *q* and *RH* for T_{dw}-0.05, T_{dw} and T_{dw} +0.05. These values are chosen to represent any potential error in T_{dw} due to original measurement as T_w or *RH* and subsequent conversion to T_{dw} where loss of decimal precision may be an issue or small errors associated with choice of conversion algorithms. The range of different *e*, *q* and *RH* values calculated for each observation are plotted as monthly and annual averages (Fig. 2.6). There is a seasonal cycle in these errors reflecting the non-linearity of humidity conversion and mean errors across the year are 0.07 hPa, 0.04 g kg^{-1} and 0.54 % for *e*, *q* and *RH* respectively.

While these errors are unavoidable, if they are random, their effect on the dataset will be substantially reduced by using monthly mean anomaly 5^{o} by 5^{o} grid-box resolution in HadCRUH. It is however, important to recognise the potential presence of these errors which should be considered further for any work in quantifying uncertainty in HadCRUH.

2.7 CONCLUSIONS

As both the land and marine data report humidity in T_{dw}, consistency of conversion algorithms can be maintained throughout HadCRUH. Of the many humidity variables, e, q and RH are selected for use in this project. Various formulae have been assessed and finally chosen as suitable for calculating e, e_s, $T_{w,}$ q and RH considering the end use for climate analyses. It is decided to use the standard atmospheric P for the marine data and an approximated P from station elevation for the land data.

2.8 TABLES AND FIGURES FOR CHAPTER 2

Table 2.1
(2Table2_1.pdf)

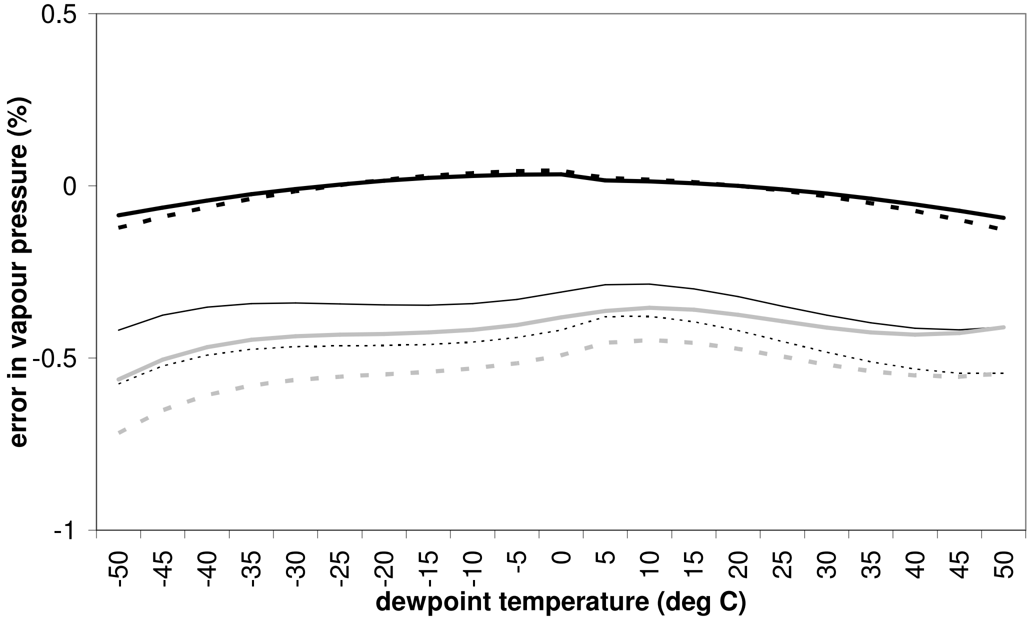

Figure 2.1: Comparison of equations for *e*. The difference series (in percentage error) between equation 2.1e (high accuracy) and equations 2.1b (thin black lines), 2.1c (thick grey lines) and 2.1d (thick black lines) are plotted. Dashed and solid lines refer to *e* calculated at atmospheric pressures of 750 hPa and 1040 hPa respectively. Note: equations 2.1b and 2.1c do not use *P* but are subtracted from timeseries created by equation 2.1e at two different levels of *P*. Equations are shown in Table 2.1.

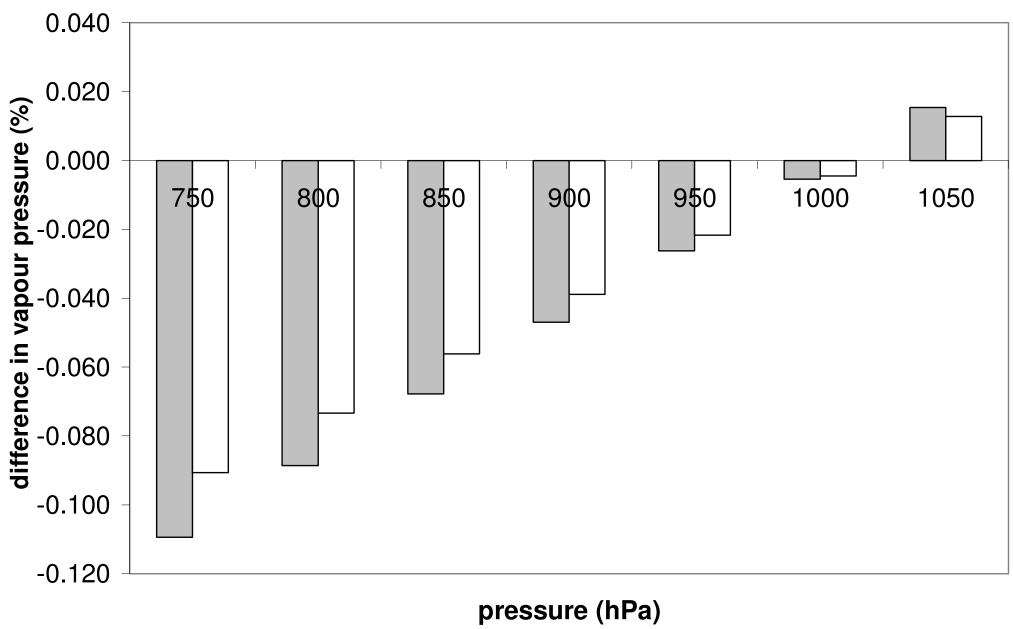

Figure 2.2: Sensitivity of *e* (calculated from equation 2.1d) to different atmospheric *P* values. Percentage differences in *e* calculated for a range of pressures compared to a standard *P* of 1013 hPa are shown for T_{dw} values below (grey bars) and above (white

bars) zero where the enhancement factor (*f*) in use is for ice-bulb and wet-bulb respectively.

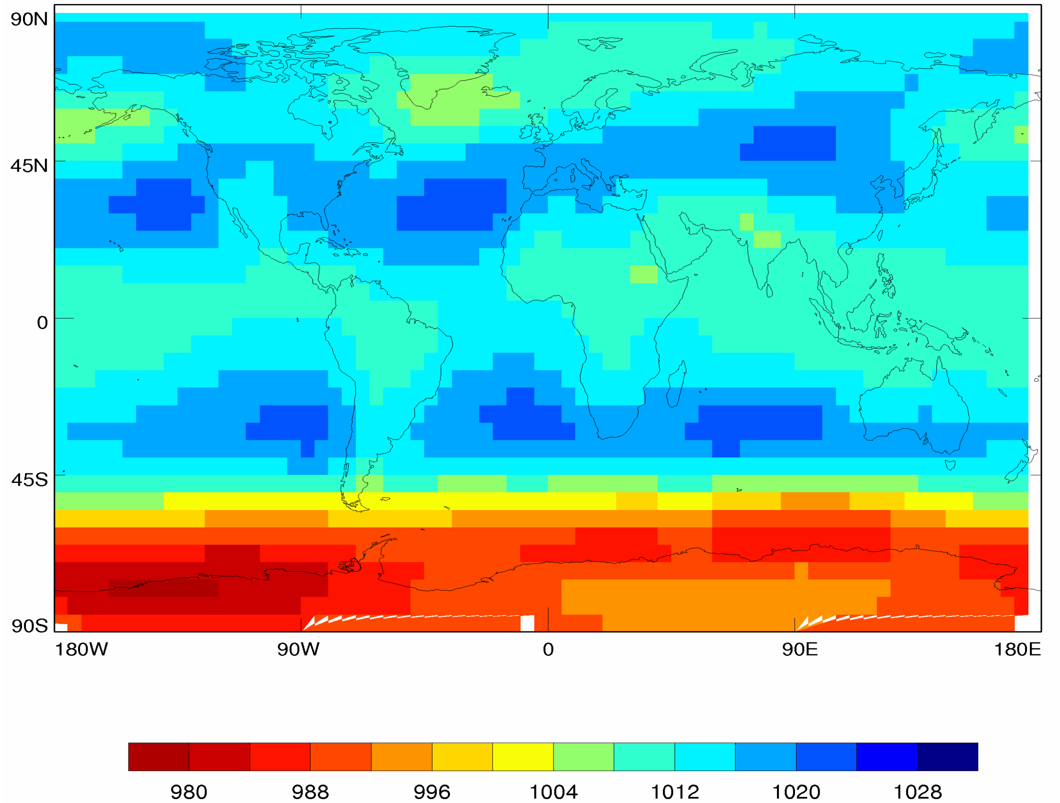

Figure 2.3: Climatological annual mean sea level pressure over the period 1974 to 2003. Data are from the global sea level pressure dataset HadSLP2 (Allan & Ansell, 2006). Units are in hPa.

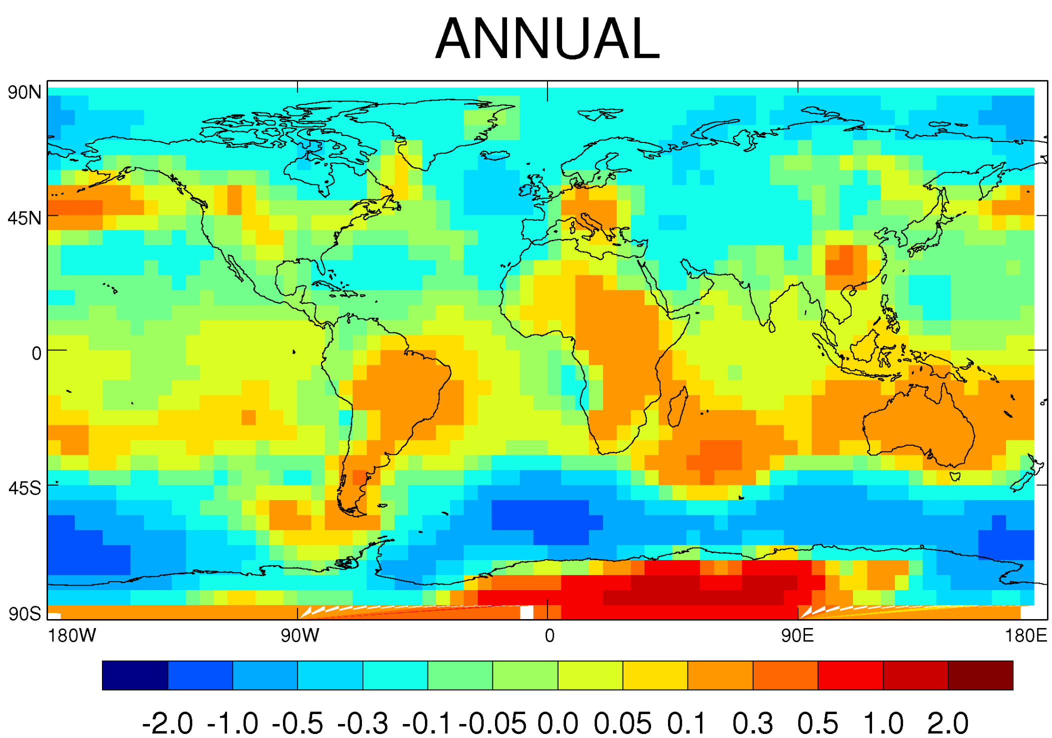

Figure 2.4: Trends in sea level pressure over the period 1973 to 2003. Units are in hPa $10yr^{-1}$. Data are from the global sea level pressure dataset HadSLP2 (Allan & Ansell, 2006). Trends are calculated using the MPS method (Box 3.4).

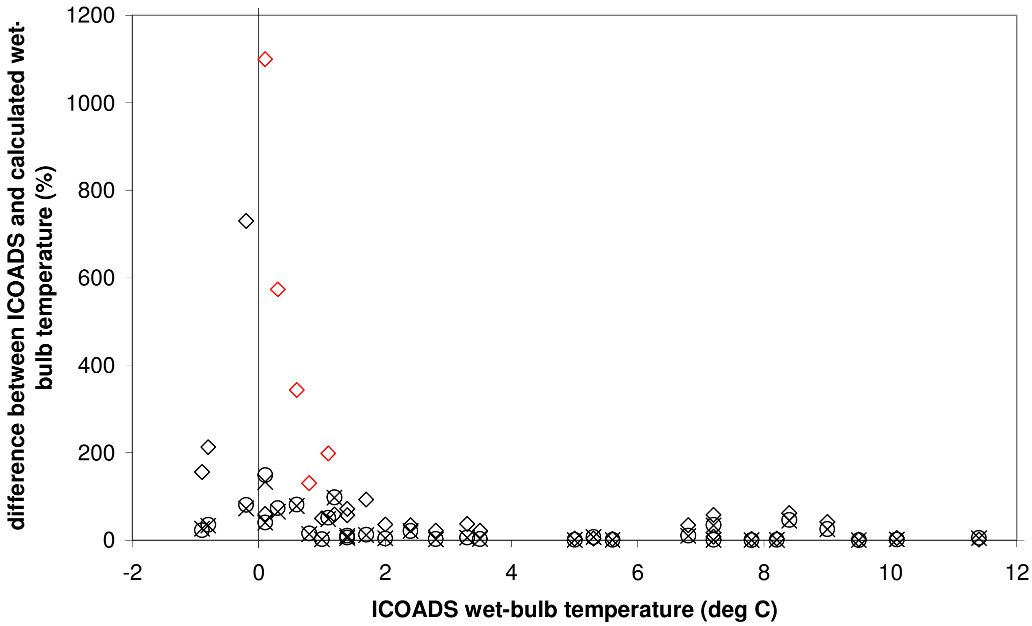

Figure 2.5: Comparison of a sample set of T_w values provided by ICOADS with calculated T_w from simultaneous T and T_{dw} values. Crosses, circles and diamonds represent equation 2.4 using the standard atmospheric P of 1013 hPa, equation 2.4 using the ICOADS reported P and equation 2.3 respectively. Black symbols show values where ICOADS T_w and calculated T_w are of the same sign and red symbols show values where the sign differs.

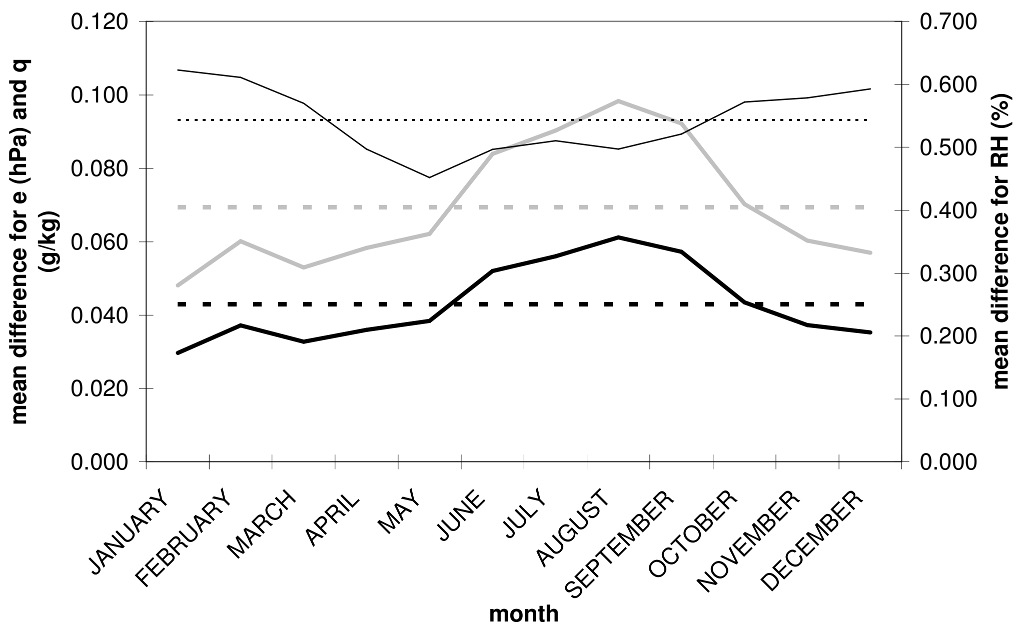

Figure 2.6: Mean error range in conversions to *e*, *q* and *RH* from errors in source variables. A sample year (1980) of hourly T and T_{dw} data from station 037720 (Heathrow, UK) is converted to *e* (grey), *q* (thick black) and *RH* (thin black) at T_{dw}-0.05 and T_{dw}+0.05 to reflect the possible errors in T_{dw} from original measurement as T_w or *RH* and subsequent conversion. Solid lines show the error range averaged monthly and

dashed lines show the error range averaged for the whole year. When analysed as ± percentage errors all are between 0.3 to 0.4 % for all three variables.