Chapter 4:
Comparison of Aircraft and Ground-based Data

Chapter Contents

4.1 Introduction

Comparisons of aircraft and ground-based observations are required to validate the calibration of airborne instruments and the processing of the data from these instruments. There is also a need to investigate the consistent under-estimation of the surface fluxes by airborne sensors identified in Shuttleworth (1991) and this discrepancy needs to be explained before airborne observations of the surface fluxes will be considered reliable. Furthermore, it is necessary to demonstrate the equivalence of aircraft and ground-based observations before these can be integrated in studies that cover multiple spatial scales. Such integration is one of the main themes of the work presented in this thesis and the comparisons presented here provide the justification for combining data from airborne and ground-based instruments.

There have been many comparisons between fluxes measured using aircraft and ground-based instruments over the last two decades. Desjardins et al. (1989) compared aircraft and ground-based measurements of the fluxes of sensible heat, latent heat and carbon dioxide. Their results show that the aircraft values of F_E and F_C were larger than the ground-based values, which they attribute to the short (2 minute) averaging period used for the ground-based measurements. In contrast, the aircraft values of F_H were less than the ground-based values, the reasons given being radiative transfer, flux divergence with height and the loss of high frequency information in the aircraft data. Kelly et al. (1992) compared aircraft measurements of F_H and F_E with ground-based data from a mixture of twenty-two eddy-covariance and Bowen ratio systems used during the 1987 FIFE experiment. Surface values of the fluxes were estimated by extrapolating aircraft measurements of flux profiles to ground level. The results show that the aircraft measurements of F_H were 20 to 50% less than the ground-based measurements in all seasons and that F_E measured by the aircraft was less than the ground-based measurements at high values of F_E and greater than the ground-based values for small F_E. The reasons put forward for the disagreement were the loss of high frequency information from the aircraft data, source-area differences between the aircraft and the ground-based systems and statistical uncertainty due to the relatively short, 15 km flight paths.

In an independent analysis of data from the 1987 and 1989 FIFE experiments, Desjardins et al. (1992a) compared data from the NRC Twin Otter with ground-based data from two eddy correlation sites. The aircraft measurements of F_H and F_E agreed with data from one site but under-estimated F_H by 40% and over-estimated F_E by 14% compared to data from a second site. The reasons for the under-estimation of F_H were given as the under-estimation of vertical velocity variance by the aircraft during the 1987 experiment and flux divergence with height. Source-area differences are also cited as a possible reason for the disagreement.

Lucotte and Saïd (1996) compared fluxes measured by aircraft with data from eight ground-based sites and found that the aircraft values for F_H were 20% less than the ground-based values with a greater under-estimation for F_E. Crawford et al. (1996) compared ground-based observations of F_H and F_E with aircraft data for the BOREAS and Alaska Landscape Flux Study (ALFS) experiments. In their case, the aircraft values for F_H under-estimated the ground-based values by 10 to 20%, the aircraft observations of F_E over-estimate the ground-based values by 25% but $F_H + F_E$ showed better agreement. Once again, the disagreement is attributed to the different source-areas of the aircraft and ground-based measurements but this is not rigorously tested.

The results from these investigations may be summarised as follows. Most comparisons find that aircraft observations of F_H under-estimate the ground-based measurements by between 10% and 50%. The results for F_E are less certain with some studies finding that aircraft observations under-estimate those from ground-based instruments, others find that the aircraft instruments over-estimate F_E and some find either depending on the magnitude of the flux. The most common reasons offered for the observed behaviour are the loss of high frequency contributions due to under-sampling, loss of low frequency contributions due to short flight paths, flux divergence with height and differences in the surface areas sampled by the aircraft and ground-based instruments. These, and other possible errors associated with both ground-based and aircraft observations, are discussed in Mann and Lenschow (1994) and Mahrt (1998). Their definitions of random and systematic error will be used here to describe, respectively, the expected scatter and bias in the comparisons due to a finite averaging length or time.

The comparison of aircraft and ground-based observations during OASIS contributes to the previous work in this area in two, new ways. First, a series of dedicated comparison flights were performed at low-level to minimise differences between the measurement heights of the aircraft and ground-based instruments. These flights were designed to allow the comparison of means and standard deviations in addition to the surface fluxes and to minimise the effect of flux divergence with height. Second, the effect of surface heterogeneity on the comparisons is estimated and then used to correct the aircraft observations of F_H and F_E. This directly tests the hypothesis that the discrepancies between aircraft and ground-based observations are attributable to differences in the source-area of the measurements.

This chapter compares aircraft and ground-based observations using data from the 1995 OASIS experiment. Section 4.2 describes the data used in the comparison and the methods used to calculate the random and systematic errors in the comparison. The results of the comparison between the aircraft and ground-based data for the low-level, grid and transect flights are presented in Section 4.3. Means of wind direction WD, wind speed U and air temperature T_a and standard deviations of potential temperature θ, wind speed, vertical velocity w and specific humidity q are compared for the low-level flights at the Wattles site. The fluxes of sensible heat F_H, latent heat F_E, net radiation F_N and the friction velocity u_* are compared for the low-level flights, the grid flights over the Browning and the Wattles sites and for the transect flights at the Wagga, Browning and Urana sites. Aircraft and ground-based measurements of the flux of CO_2, F_C, are compared for the grid and transect flights (Wagga and Browning only), there were no ground-based measurements of F_C at the Wattles or Urana sites. The comparison results are discussed in Section 4.4 and conclusions presented in Section 4.5.

4.2 Data and Techniques

4.2.1 General

Data from three sources within the 1995 OASIS field experiment are used to compare measurements of meteorological quantities from the aircraft and from ground-based instruments.

The first data set comes from the low-level flights at approximately 6 m over two instrumented fields and the adjoining non-instrumented fields at the Wattles site. The low-level of these flights was chosen to allow the direct comparison of means and standard deviations from the aircraft and ground-based instruments and to minimise any differences between the aircraft and ground-based measurements due to differences in their respective footprints and the divergence of the fluxes with height. Quantities compared at this site were U, WD, T_a, standard deviations of vertical velocity, σ_w, horizontal wind speed, σ_u, temperature, σ_θ and specific humidity, σ_q, fluxes of sensible heat, latent heat and net radiation and the friction velocity.

The second data set comes from a series of grid flights made at heights between 15 and 65 m and which overflew three instrumented fields at the Browning site (oats, pasture and canola) and the instrumented fields at the Wattles site. The grid flights were intended to provide observations of spatially averaged fluxes and comparisons that use the data from these flights will be subject to errors due to flux divergence with height and differences in source-area. Means and standard deviations were not compared using this data set because of the height difference between the aircraft and the ground-based measurements. Quantities compared using data from the grid flights were F_H, F_E, $S_\downarrow$, F_N, u_* and F_C.

The third data set comes from transects flown at approximately 20 m between the Wagga and Urana sites. The aircraft passed within 1 km of the Wagga site and within 500 m of the Browning and Urana sites during the transects. As with the grid flights, the aircraft height above ground during the transect flights restricts the comparison to F_H, F_E, $S_{\downarrow}$, F_N, u_* and F_C.

4.2.2 Data Processing

Turbulence data measured by the aircraft instruments had linear trends and means removed prior to calculating the fluxes of sensible heat, latent heat, CO_2 and momentum. Removal of the mean vertical wind speed from the data for each run restricts the longest wavelength contributing to the flux to the length of the aircraft run. This corresponds to about 1200 m for aircraft data collected during comparisons at the Wattles site and about 10,000 m for aircraft data collected during the grid and the transect flights. The loss of long wavelength contribution to the fluxes is discussed in Section 4.4.3, loss of short wavelength contributions is discussed in Section 4.2.4.

The contamination of the aircraft data by virtue of movement through vertical gradients has also been considered. There are two mechanisms by which this may occur. First, the aircraft height above ground level can change even when the aircraft is flying over flat terrain. No relationship was found between the variance of the aircraft height above ground level and the scalar variances or the correlation between the scalars and the vertical velocity. This indicates that any contamination must be small. Second, the aircraft may not be able to follow all terrain height variations. However, the variance of the aircraft height above sea level is generally larger than the variance of the aircraft height above ground level as expected when the aircraft is able to follow the terrain. This suggests that any contamination of the time series data due to changes in aircraft height above ground level forced by rapid changes in terrain height above sea level is small.

The ground-based measurements of means and standard deviations used in the comparison at the Wattles came from instruments mounted on the mast in the wheat field. Fluxes were averages of the values from the wheat and oat fields for east-west aircraft passes and from the wheat field alone for north/south aircraft passes. When the aircraft flight spanned two 30 minute averaging periods, the final ground-based value was calculated as a weighted average of the data from both periods with the weights determined by the fraction of each period covered by the aircraft flight.

Ground-based measurements of the fluxes used in the comparison at the Browning site were averages of the data from the three sites (pasture, oats and canola) at Browning and the data from the wheat field at the Wattles site. All ground-based data were linearly interpolated onto a 1 minute time step and then averaged over all available sites before extracting the period covered by the aircraft grid flight and averaging this to produce a single ground-based value for the time of the aircraft flight.

4.2.3 Random Errors in Variance and Covariance Measurements

Mahrt (1998) describes several sources of error in aircraft and ground-based measurements of variances and covariances. The divergence of vertical fluxes with height, loss of long wavelength contributions due to finite averaging times and surface heterogeneity are discussed in Section 4.4. This section considers the random error expected in measurements of variance and covariance by both aircraft and ground-based systems due to the finite averaging length or time. This provides a means for assessing the scatter in the comparisons presented.

The random errors in the variance and covariance are estimated using the following expressions from Dobosy et al. (1997):

$$\begin{aligned} \varepsilon_{s^2}^2 &= \sigma_s^2 \frac{2L_s}{UP_{AV}} \\ \varepsilon_{ws}^2 &= F_s^2 \frac{2L_{ws}}{UP_{AV}} \left[\frac{1+r_{ws}^2}{r_{ws}^2} \right] \end{aligned} \qquad 4.1$$

where $\varepsilon_{s^2}^2$ and ε_{ws}^2 are the random errors in the variance and covariance measurements respectively. Here, the quantity s can be the vertical velocity, specific humidity, potential temperature or wind speed. In Equation 4.1, σ_s^2 is the ensemble variance of s, F_s is the vertical flux of s, L_s is the integral length scale of s, L_{ws} is the integral length scale of the covariance of w and s, r_{ws} is the correlation coefficient between w and s, U is the wind speed and P_{AV} is the averaging period. When calculating the random error for the aircraft data, U is taken as the aircraft airspeed and P_{AV} as the duration of the aircraft run. The total random error in the difference between standard deviations measured by the aircraft and ground-based instruments is taken as:

$$SD_{ran} = \frac{1}{2}\left(\varepsilon_{AC}^2 + \varepsilon_{GB}^2\right)^{1/2} \qquad 4.2$$

where the subscripts AC and GB refer to the aircraft and ground-based measurements respectively and ε^2 can be either $\varepsilon_{s^2}^2$ or ε_{ws}^2 given by Equation 4.1.

Estimates of L_s and L_{ws} for Equation 4.1 were obtained by fitting the cospectral and spectral forms given in Kaimal and Finnigan (1994) to the observed cospectra of $w\theta$, wq and wu and the spectra of w, θ and q (all heights) and u (6 m only). The length scales for θ and q and for $w\theta$, wq and wu were the same to within the uncertainty of the method. Spectra of u for the grid flights did not show a clearly defined peak due to the relatively short length of the grid legs. The integral length scale of u for these flights was estimated from the lag at which the autocorrelation function dropped to a value of $1/e \approx 0.37$ (Kaimal and Finnigan, 1994). Length scales estimated by this method were similar to those obtained from the spectral fit method for aircraft data recorded at 6 m.

The aircraft measurements of mean quantities used in the comparison at the Wattles site were calculated by averaging the values from each low-level pass. Standard deviations were calculated from the mean of the pass variances plus the variance of the pass means and fluxes were calculated by averaging the fluxes for each pass. Data for F_H and F_E from individual passes were only included in the average if they satisfied the condition $0.5 \le \left(F_H + F_E\right)/\left(F_N - F_G\right) \le 1.5$. This was done to exclude data from passes that did not approximately close the surface energy budget so as to prevent contamination of the flight average. The same process was used to produce mean fluxes for the grid flights by averaging the values from each grid leg. Averaging the values from each low-level pass or grid leg reduces the random error in the final figure by a factor of $1/\sqrt{N}$ where N is the number of passes (ten to fourteen) or grid legs (six). In contrast, the aircraft fluxes used in comparisons involving data from the transect flights are single values taken from the transect segment nearest to the ground-based site. This means that random errors are expected to be larger when using data from transect flights than when using data from the low-level and grid flights.

In addition to the expected random error calculated above, the scatter between variance and covariance measurements during the 1976 International Turbulence Comparison Experiment (ITCE76, Dyer et al., 1982) is used to assess the scatter in the comparisons presented here.

4.2.4 Spectra and Length Scales

The effect of the aircraft temperature sensor response time on the measurements of σ_θ and F_H was discussed in Section 3.7.2 in Chapter Three. The loss of high frequency contributions to the aircraft observations of $w\theta$ and wq due to phase differences between w and the scalars θ and q was discussed in Section 3.7.5 in Chapter Three. This section examines the loss of high frequency contributions to the aircraft measurements of scalar variances due to under-sampling and presents results for the length scales and correlation coefficients needed to estimate the random error in the aircraft and ground-based observations.

Figure 4.1 shows the average spectra for θ and q for the low-level flights at the Wattles site. The spectral forms given in Kaimal and Finnigan (1994) are also plotted. The θ and q spectra at 6 m show no contributions to the variance above the Nyquist frequency $n = fz/u = 10 \times 6/40 = 1.5$ and the same effect is also present in the aircraft w and u data (not shown). The consequent loss of variance, estimated from the Kaimal and Finnigan (1994) spectral forms, is 18% and the aircraft measurements of σ_θ, σ_q, σ_w and σ_u at 6 m have been multiplied by a factor of 1.09 to correct for this loss.

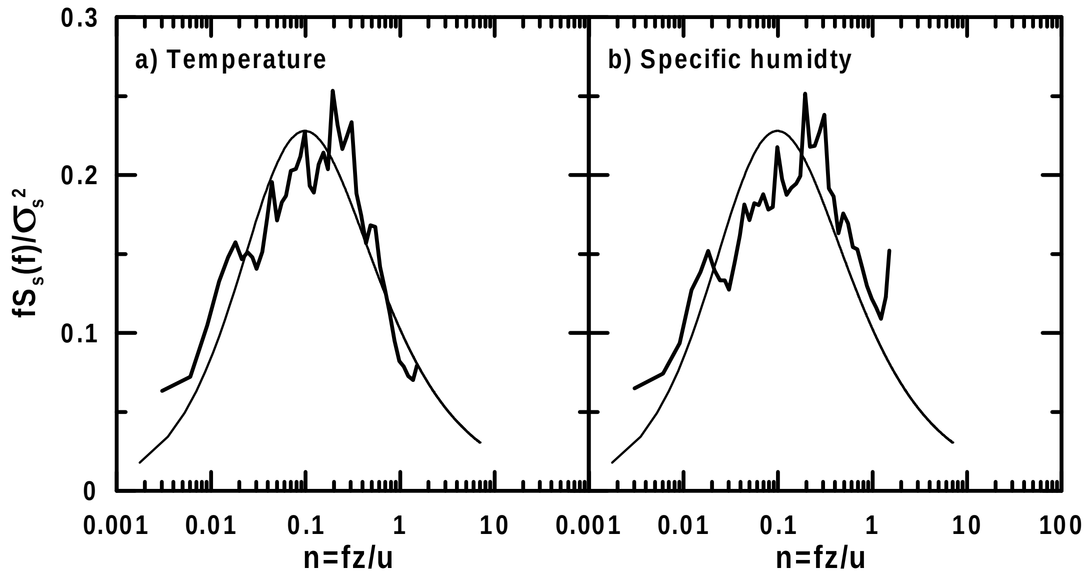

Figure 4.1 Average spectra of a) θ and b) q for aircraft data during the low-level flights at the Wattles site (thick line). The spectral form from Kaimal and Finnigan (1994) is plotted as thin lines.

Table 4.1 lists the length scales for w, θ, u and $w\theta$ and the correlation coefficients for $w\theta$, wq and wu for the low-level (6 m) and grid (average height 35 m) flights. The length scales at 35 m are approximately 20% smaller than those predicted using the formulae in Lenschow and Stankov (1986) but are a factor of 3 to 4 smaller at 6 m. This reflects the fact that the Lenschow and Stankov (1986) formulae are derived from aircraft flights in the mixed layer and predict that the height dependency of the length scales is $z_i z^{1/2}$ where z_i is the mixing height. Close to the surface during the low-level flights, the length scales are approximately equal to the measurement height, a finding confirmed by these results. Length scales and correlation coefficients calculated from the aircraft data at 35 m are similar to those quoted in Dobosy et al., (1997). The loss of correlation between w and u with height occurs because 10 of the 13 aircraft grid flights considered here took place above the surface layer ($z/L < -1$, Kaimal and Finnigan, 1994).

Table 4.1 Integral length scales for vertical velocity, L_w, temperature, L_θ, horizontal wind speed, L_u and temperature flux, $L_{w\theta}$, and the correlation coefficient between vertical velocity and temperature, $r_{w\theta}$, specific humidity, r_{wq} and horizontal wind speed, r_{wu}, for aircraft flights at 6 and 35 m.

z	L_w	L_θ	L_u	$L_{w\theta}$	$r_{w\theta}$	r_{wq}	r_{wu}
m	m	m	m	m			
6	8 ± 2	7 ± 3	50 ± 40	10 ± 1	0.51	0.43	-0.18
35	40 ± 20	55 ± 10	280 ± 80	55 ± 10	0.55	0.48	-0.09

4.3 Results of Comparisons

4.3.1 Low-level Flight Comparisons

Figure 4.2 compares the aircraft and the ground-based observations recorded during the low-level flights at the Wattles site. The statistics of the comparison are given in Table 4.2. The site layout and the aircraft flight tracks are shown in Figure 2.2 in Chapter 2.

The aircraft and ground-based observations of mean wind direction WD, wind speed U and air temperature T_a (Figure 4.2a, b and c) are in close agreement with small mean differences of 4 °, 0.2 m s^{-1}, and 0.9 °C respectively. The good agreement between the aircraft and ground-based observations of the mean values confirms the calibrations derived in Chapter Three.

Aircraft and ground-based measurements of σ_w, σ_u, σ_θ and σ_q are compared in Figure 4.2d, e, f and g. The slope of the best-fit line through the origin is close to one except for σ_θ where the aircraft measurements under-estimate the ground-based observations by 8%. The correlation between aircraft and ground-based measurements is high for σ_w, σ_u and σ_θ but somewhat lower for σ_q. The expected random error, SD_{ran} (Equation 4.2), accounts for between 20% to 30% of the scatter in the w, θ and q comparisons and is similar to the observed scatter for the u comparison. The scatter in the results for σ_q is similar to that found during ITCE76 but the scatter in σ_w, σ_θ and σ_u is twice as large as the ITCE76 results. The good agreement between the aircraft and ground-based observations of the variances is further confirmation of the calibration of the aircraft sensors. The agreement between the ground-based and aircraft observations of σ_w and σ_u is also evidence that the aircraft motion is correctly measured and the effect of this on the measurement of the turbulent wind field is correctly removed during the data processing. This follows because both under- and over-compensation for the aircraft motion will increase the variance in the calculated wind field, resulting in disagreement between the aircraft and ground-based observations.

The aircraft and ground-based measurements of u_*, F_N, F_H and F_E are compared in Figure 4.2h, i, j and k respectively. The open circle in Figure 4.2h corresponds to the low-level flight with the worst contamination of dynamic pressure (see Chapter Three) and this point has been excluded from the regression statistics. Aircraft and ground-based observations of u_* compare well but aircraft observations of F_N are 20% larger than ground-based data. F_H derived from the aircraft data under-estimates the ground-based values by 13% but this is a substantial improvement over the factor of two under-estimation reported in Isaac and McAneney (1997). For F_E, the aircraft measurements overestimated the flux by 11% with respect to the ground-based system.

The expected random error accounts for between 30% and 60% of the scatter observed in the comparison of F_H and F_E at the Wattles site. The largest contribution to the random error comes from the 30-minute averaging period used for the ground-based data. For u_*, the observed scatter is somewhat smaller than the expected random error. For F_H, F_E and u_*, the observed scatter is comparable to, or less than, the scatter found in the ITCE76 results. The good agreement between the aircraft and ground-based observations of the fluxes is further confirmation of the calibration of the aircraft sensors and validation of the data processing. Comparison of the observed and expected random error shows that much of the scatter can be attributed to the intermittent nature of the transport processes and, in addition, the observed scatter is similar to that found during a careful ground-based inter-comparison of turbulence instrumentation.

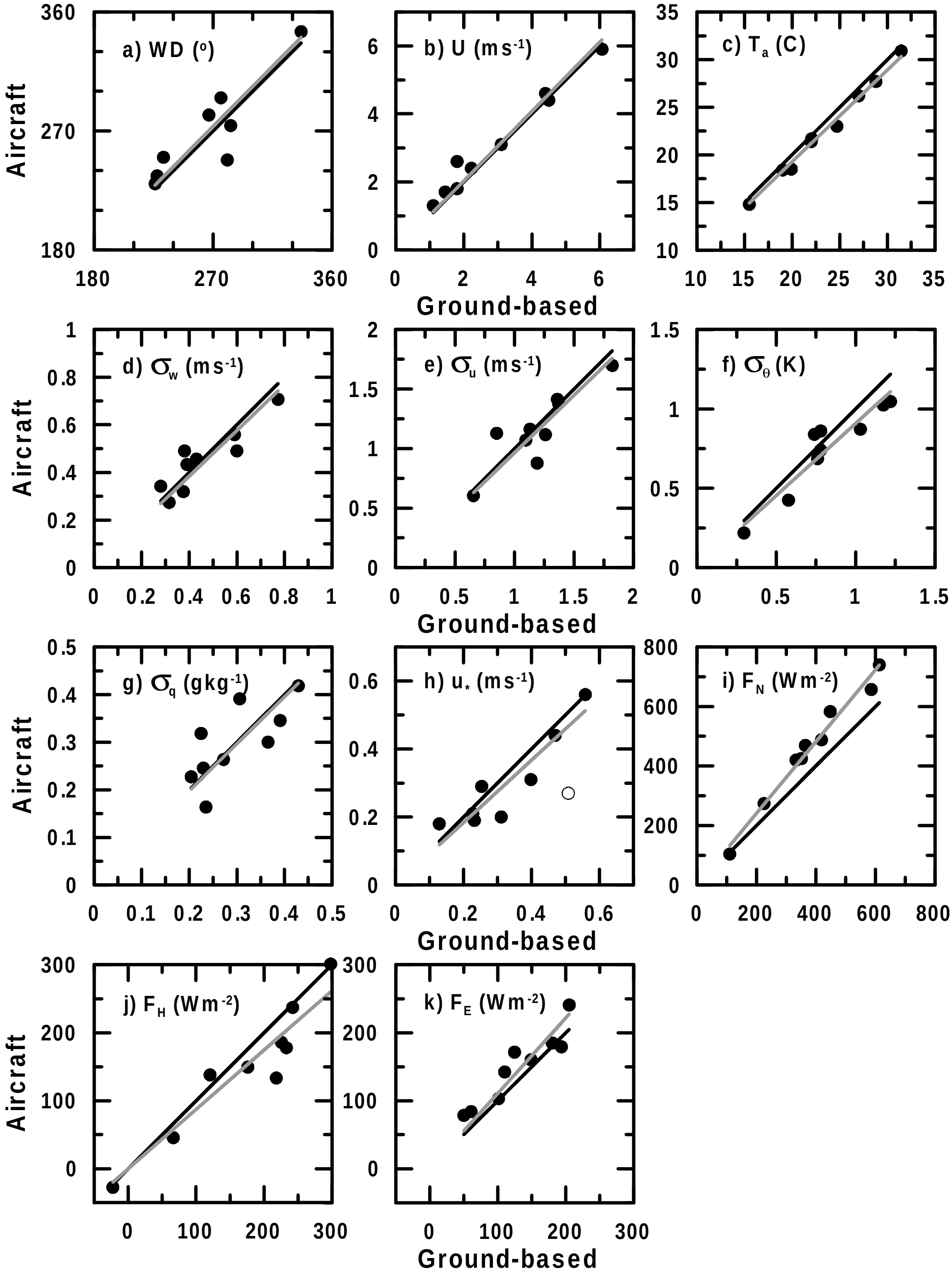

Figure 4.2 Comparison of aircraft and ground-based measurements of mean a) wind direction WD, b) wind speed U and c) air temperature T_a, standard deviation of d) vertical velocity σ_w, e) wind speed σ_u, f) potential temperature σ_θ and g) specific humidity σ_q and h) friction velocity u_*, i) net radiation F_N, j) sensible heat flux F_H and k) latent heat flux F_E at the Wattles site. The black line is 1:1 and the grey line is the line of best fit.

Table 4.2 Comparison of aircraft and ground-based observations at the Wattles site. m is the slope of the best fit line through the origin, r^2 is the correlation coefficient squared, MD is the mean difference (ground-based minus aircraft) and SD is the standard deviation of the difference expressed as a percentage of the mean of the ground-based data. SD_{ran} is the estimated random error in the difference expressed as a percentage of the mean of the ground-based data and SD_{ITCE76} is taken from Dyer et al. (1982). SD_{ran} and SD_{ITCE76} are not calculated for means and F_N. The uncertainty for m is the 90% confidence interval.

	m	r^2	MD		SD %	SD_{ran} %	SD_{ITCE76} %
WD	1.01 ± 0.03	0.79	-4	º	6	-	-
U	1.02 ± 0.04	0.96	-0.2	m s^{-1}	10	-	-
T_a	0.96 ± 0.01	0.99	0.9	C	2	-	-
σ_w	0.96 ± 0.07	0.73	0.01	m s^{-1}	15	4	8
σ_u	0.97 ± 0.06	0.75	0.17	m s^{-1}	13	10	6
σ_θ	0.92 ± 0.05	0.87	0.07	K	12	4	8
σ_q	0.99 ± 0.09	0.46	0.00	g kg^{-1}	20	4	19
u_*	0.92 ± 0.08	0.85	0.05	m s^{-1}	26	20	34
F_N	1.20 ± 0.03	0.98	-79	W m^{-2}	11	-	-
F_H	0.87 ± 0.07	0.90	24	W m^{-2}	18	6	16
F_E	1.11 ± 0.07	0.82	-16	W m^{-2}	15	9	34

4.3.2 Grid Flight Comparisons

Figure 4.3 compares the aircraft and ground-based measurements of u_*, $S_\downarrow$, F_N, F_H, F_E and F_C for the grid flights and the statistics for this comparison are given in Table 4.3. The difference in measurement height between the aircraft and ground-based systems prevents the comparison of means and variances for these flights. The location of the ground-based sites and the aircraft flight track are shown in Figure 2.3 in Chapter 2.

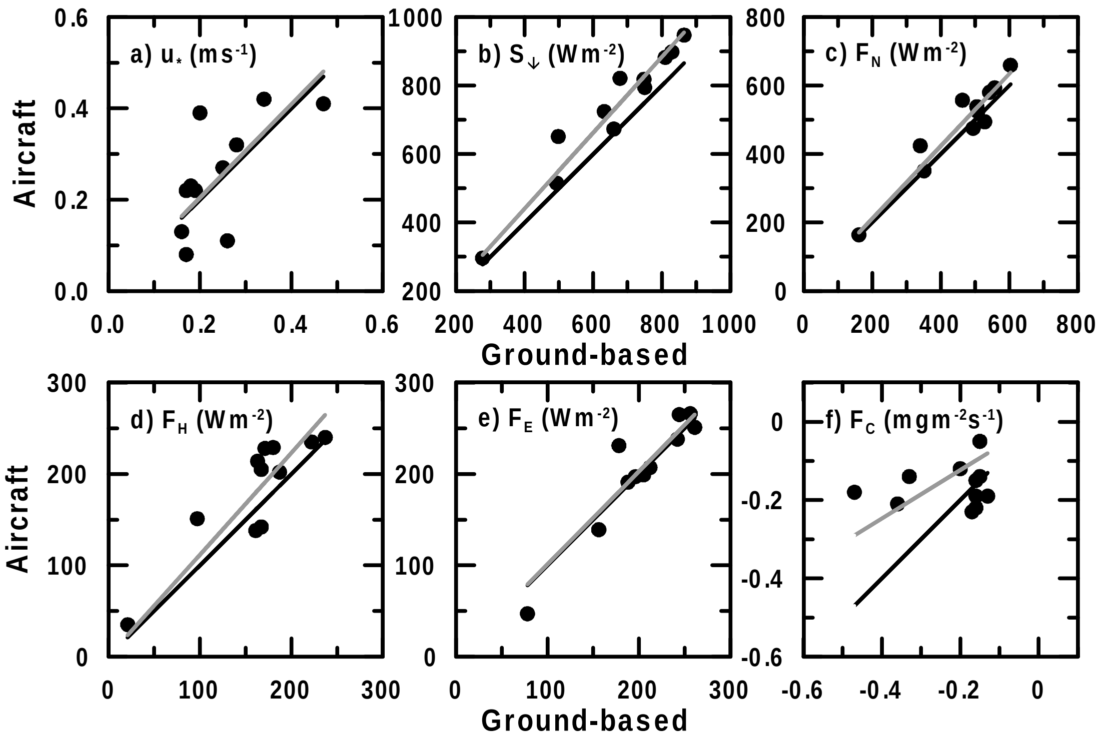

Figure 4.3 Comparison of aircraft and ground-based measurements of the fluxes of a) friction velocity u_*, b) incoming shortwave radiation $S_\downarrow$, c) net radiation F_N, d) sensible heat F_H, e) latent heat F_E and f) CO_2 flux F_C for the grid flights. The black line is 1:1 and the grey line is the line of best fit through the origin.

Table 4.3 Comparison of u_*, $S_\downarrow$, F_N, F_H, F_E and F_C for the grid flights. Columns are as for Table 4.2. SD_{ran} is not calculated for $S_\downarrow$ and F_N. SD_{ITCE76} is not available for $S_\downarrow$, F_N and F_C.

	m	r^2	MD		SD %	SD_{ran} %	SD_{ITCE76} %
u_*	1.02 ± 0.15	0.41	-0.01	m s^{-1}	37	40	34
$S_\downarrow$	1.10 ± 0.03	0.94	-71	W m^{-2}	7	-	-
F_N	1.06 ± 0.04	0.91	-28	W m^{-2}	9	-	-
F_H	1.12 ± 0.06	0.74	-22	W m^{-2}	18	5	16
F_E	1.01 ± 0.04	0.89	-1	W m^{-2}	11	7	34
F_C	0.68 ± 0.19	0.00	-0.05	mg m^{-2} s^{-1}	60	20	-

Aircraft observations of $S_\downarrow$ and F_N are 10% and 6% larger than the ground-based measurements respectively. F_H measured by the airborne instruments overestimates the ground-based values by 12% but there is excellent agreement between the two measurement approaches for F_E. The aircraft observations of F_C are 32% smaller than the ground-based values. The expected random error accounts for 28% of the observed scatter in the comparison of F_H and 64% of the observed scatter in the comparison of F_E. The observed scatter in the F_H comparison is similar to that found during ITCE76 while that for the F_E comparison is considerably smaller.

There was poor agreement between the aircraft and ground-based measurements of u_* ($r^2 = 0.41$) compared to the results from the low-level flights. The most likely explanation is that during the low-level flights the aircraft was within the surface layer and, in common with the ground-based system, was able to measure the surface stress. However the grid flights were performed at between 15 and 65 m with $|z/L| > 1$ for 10 of the 13 flights, so that the aircraft was in the free convection layer (Kaimal and Finnigan, 1994) for most grid flights. Here, the correlation between w and u breaks down, u_* is no longer the appropriate velocity scale, and the aircraft measures a local stress, not the surface stress measured by the ground-based instruments.

4.3.3 Transect Flight Comparisons

Figure 4.4 shows the comparison between aircraft and ground-based measurements of u_*, $S_\downarrow$, F_N, F_H, F_E and F_C. The comparison statistics are given in Table 4.4. The locations of the ground-based sites and the aircraft flight track are shown in Figure 2.4 in Chapter 2.

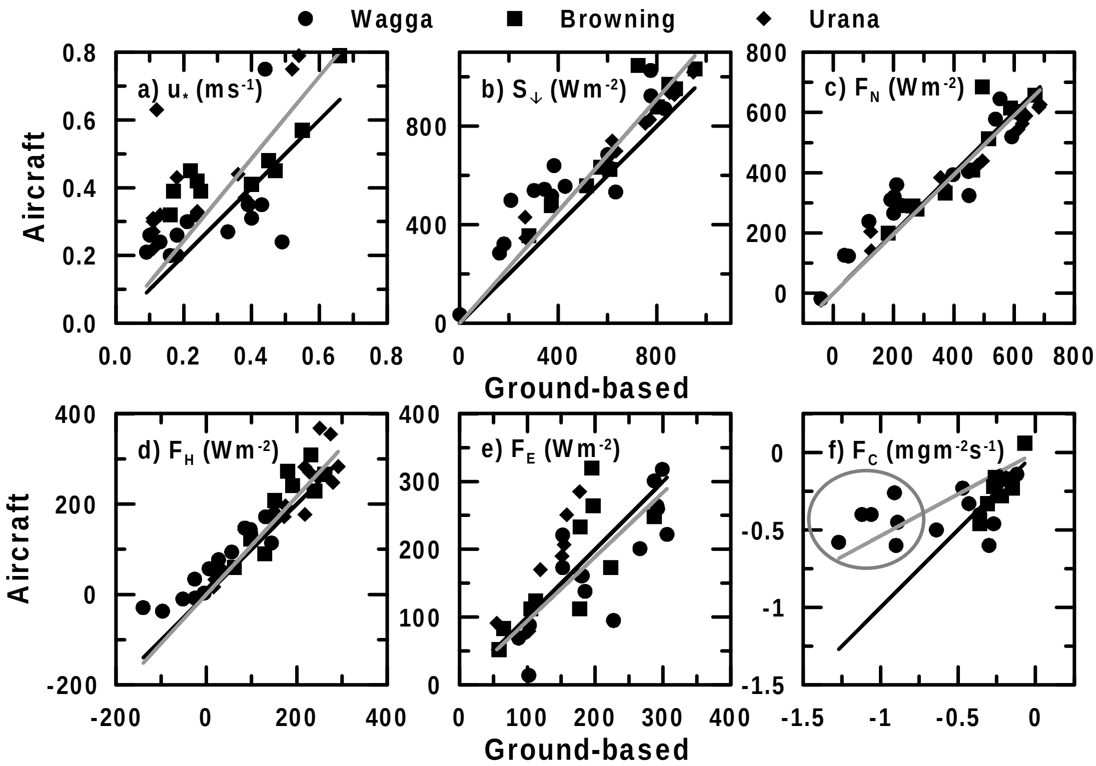

Figure 4.4 Comparison of aircraft and ground-based measurements of the fluxes of a) friction velocity u_*, b) incoming shortwave radiation $S_\downarrow$, c) net radiation F_N, d) sensible heat F_H, e) latent heat F_E and f) CO_2 flux F_C for the transect flights. The black line is 1:1 and the grey line is the line of best fit through the origin.

Table 4.4 Comparison of u_*, $S_\downarrow$, F_N, F_H, F_E and F_C at the transect sites. Columns are as for Table 4.3. SD_{ran} is not calculated for $S_\downarrow$ and F_N. SD_{ITCE76} is not available for $S_\downarrow$, F_N and F_C.

	m	r^2	MD		SD %	SD_{ran} %	SD_{ITCE76} %
u_*	1.21 ± 0.11	0.11	-0.11	m s^{-1}	46	100	34
$S_\downarrow$	1.14 ± 0.04	0.84	-107	W m^{-2}	14	-	-
F_N	0.98 ± 0.04	0.83	-13	W m^{-2}	19	-	-
F_H	1.08 ± 0.07	0.82	-29	W m^{-2}	34	23	16
F_E	0.95 ± 0.07	0.62	5	W m^{-2}	31	25	34
F_C	0.54 ± 0.09	0.00	-0.21	mg m^{-2} s^{-1}	59	43	-

The comparison of aircraft and ground-based measurements of u_* shows a bias of 21%, aircraft larger than ground-based and has poor correlation. This is similar to the result found for the grid flight comparisons and emphasises the fact that the surface stress can not be estimated from aircraft at heights $|z/L|>1$. Aircraft observations of $S_{\downarrow}$ are 14% larger than the ground-based values but there is no significant bias in the comparison of F_N. The correlation between aircraft and ground-based values is high in both cases. The aircraft measurements of F_H are 8 % larger than the ground-based values and the two observations show a high degree of correlation. The expected random error accounts for 68% of the observed standard deviation in the difference between the two measurements. In contrast, aircraft measurements of F_E are 5 % smaller than the ground-based values, the comparison shows less correlation than for F_H and the observed scatter is 25% larger than the expected random error.

The aircraft measurements of CO_2 flux under-estimate the ground-based values by almost a factor of two. Six points from Wagga, circled in Figure 4.4f, dominate this comparison and if they are omitted, the slope of the best fit line becomes 0.89. The six points occur on days of strong warm air advection, when the sensible heat flux at Wagga was negative, or on days immediately following rainfall. The corresponding ground-based latent heat fluxes at Wagga were between 10% and 36% larger than the aircraft values on these days but no significant bias is found at Browning or Urana. These results are consistent with the Wagga pasture and triticale showing greater vigour than the surrounding fields covered by the aircraft. Chapter Six describes the calculation of the *NDVI* of the source-areas influencing the ground-based and aircraft observations. The values for the Wagga triticale and pasture fields are 0.86 and 0.77 respectively, both larger than the value of 0.75 for the 10 km segment of the aircraft transect adjacent to the Wagga tower. Chapter Seven presents the results of an analysis of the daily averaged fluxes that shows the bias between the aircraft and ground-based F_C is reduced to 30% when the aircraft data are scaled to the average *NDVI* of the Wagga site. This suggests that the outlying points in Figure 4.4f are due to differences in the source-areas of the aircraft and ground-based observations and not to systematic errors in the aircraft measurement of F_C.

4.3.4 The Surface Energy Budget

Closure of the surface energy budget (Equation 1.2, Chapter One) can be used as a check of the independently measured components, provided an estimate of the soil heat flux is available. F_G was measured at the ground-based sites during the 1995 OASIS experiment. For the aircraft data, F_G can be estimated as a linear function of net radiation with the functions determined by least-square regression of the ground-based measurements of F_G and F_N for the times of the comparison. Also, comparison of ground-based and aircraft observations of the sum of the turbulent heat fluxes, $F_H + F_E$, can be used to assess the degree to which surface heterogeneity is responsible for discrepancies in the individual comparisons of F_H and F_E. Local variations in the partitioning of $F_A = F_N - F_G$ will cause bias in the comparisons of F_H and F_E because of the different source-areas of the aircraft and ground-based measurements but the sum of the fluxes, $F_H + F_E$, is less sensitive to the partitioning of F_A. Bias in the comparisons of F_H and F_E but better agreement in the comparison of $F_H + F_E$ would suggest surface heterogeneity as a possible cause of the differences between aircraft and ground-based observations of the fluxes.

Figure 4.5a and b show the comparison of $F_N - F_G$ and $F_H + F_E$ measured by the ground-based and airborne instruments. The data from the low-level (9), grid (10) and transect (27) flights have been combined for this comparison. Figure 4.5c, d and e shows the comparison of ground-based and aircraft measurements of $F_H + F_E$ for the low-level, grid and transect flights respectively. The comparison statistics are listed in Table 4.5.

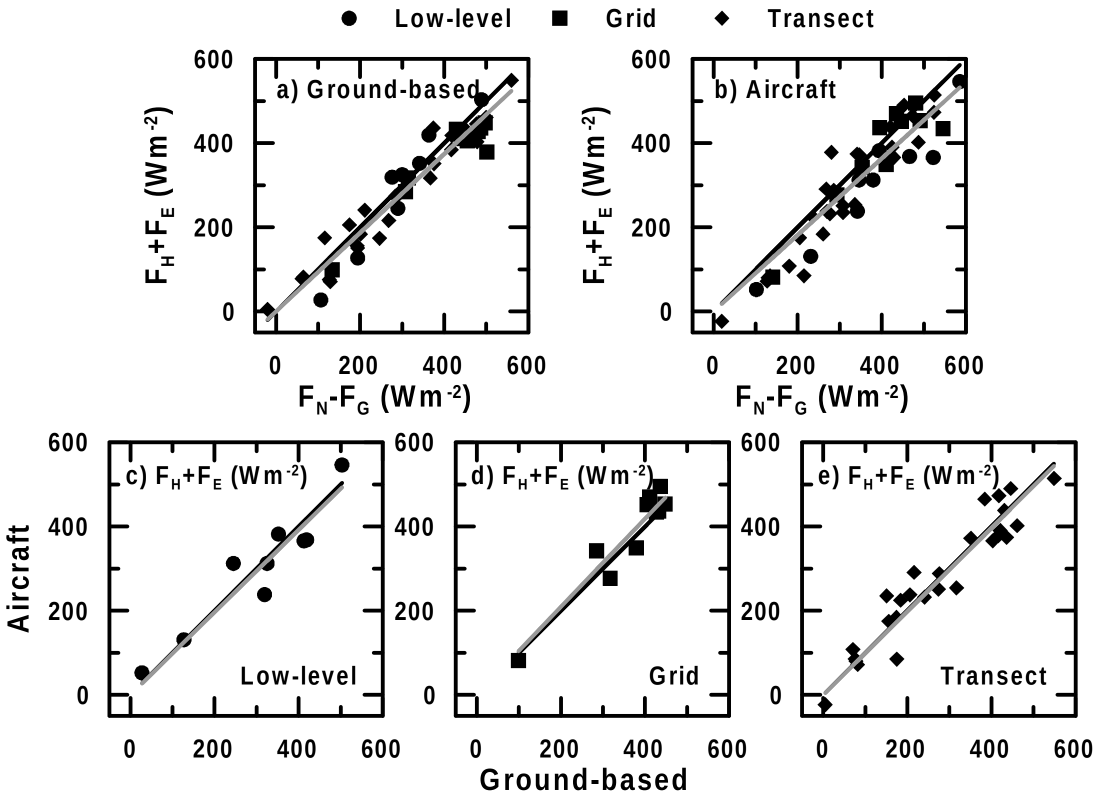

Figure 4.5 Comparison of $F_H + F_E$ and $F_N - F_G$ for a) the ground-based and b) the aircraft data and comparison of the ground-based and aircraft observations of $F_H + F_E$ for the c) low-level, d) grid and e) transect flights. The black lines are the 1:1 lines and the grey lines are the lines of best fit through the origin.

Table 4.5 Statistics for surface energy budget closure for the ground-based data, SEB closure for the aircraft data and for the comparison of ground-based and aircraft observations of $F_H + F_E$. Columns are as for Table 4.2.

		m	r^2	MD W m^{-2}	SD %
SEB	Ground-based	0.94 ± 0.02	0.93	21	13
	Aircraft	0.91 ± 0.03	0.87	37	17
$F_H + F_E$	Low-level	0.98 ± 0.07	0.89	2	16
	Grid	1.05 ± 0.04	0.91	-15	10
	Transect	0.99 ± 0.04	0.90	-4	16

The ground-based and aircraft observations of $F_H + F_E$ are smaller than $F_N - F_G$ by 6% and 9% respectively. The disparity between $F_N - F_G$ and $F_H + F_E$ measured by eddy covariance systems has been consistently observed and possible reasons are discussed in Lee (1998), Paw U et al. (2000) and Finnigan et al. (2003). The reasons for the small bias have not been explored in this work. Both ground-based and aircraft observations of F_H and F_E have been adjusted to force closure of the surface energy budget in subsequent analyses, see Chapter Five. In the present context, it is sufficient to note that the difference in the degree of under-estimation by the ground-based and aircraft systems is not statistically significant and that the two data sets show similar levels of correlation ($r^2 = 0.93$ and 0.87) and similar scatter ($SD = 13\%$ and 17%). The results demonstrate that the performance of the aircraft observations in closing the SEB is only marginally worse than that of the ground-based observations and not significantly inferior as found by Shuttleworth (1991).

The comparison of ground-based and aircraft observations of $F_H + F_E$ shows no significant bias, in contrast to the significant bias in the individual comparisons of F_H and F_E presented in Sections 4.3.1, 4.3.2 and 4.3.3. Together, the near closure of the SEB, the bias in the individual comparisons of F_H and F_E and the lack of bias in the comparison of $F_H + F_E$ strongly suggest that surface heterogeneity is responsible for the discrepancy between aircraft and ground-based observations of F_H and F_E. This and other possible sources of systematic error are discussed in the following section.

4.4 Discussion of Systematic Errors

Possible explanations for the differences between ground-based and aircraft observations of the radiative and turbulent fluxes are instrumental errors, flux divergence with height, loss of high or low frequency contributions to the fluxes and differences between the respective source-areas. This section discusses these sources of systematic error.

4.4.1 Instrumental Errors

The comparison of data from the low-level, grid and transect flights combined shows that the aircraft observations of $S_\downarrow$ and F_N are larger than the ground-based measurements by 10% and 4% respectively. The overestimate of $S_\downarrow$ by the aircraft pyranometer was assumed to be due to a calibration error and all aircraft $S_\downarrow$ data have been reduced by 10% in subsequent analyses. This also removes the bias in the comparison of F_N except for the low-level flights at the Wattles site where the aircraft observations remain 5% larger than the ground-based measurements.

The comparison of aircraft and ground-based observations of u_* using all data shows a bias of 11% but the correlation between aircraft and ground-based measurements is low for the grid and transect flights when the aircraft is often above the surface layer. The implications of this on the calculation of quantities that depend on u_* are discussed in Chapter Five. The under-estimation of F_C by the aircraft sensors may be due to either instrumental errors or differences in the source-areas of the aircraft and ground-based measurements. The lack of a reliable fast response CO_2 sensor on the aircraft and the lack of F_C observations at the Wattles site, where source-area differences were minimised, prevent this discrepancy from being resolved. The aircraft measurements of F_C are used without further correction in subsequent analyses.

The ground-based sites used accepted micrometeorological sensors and techniques and systematic errors in F_H and F_E due to instrumental effects are expected to be small in these data. Problems with the aircraft instruments were discussed in Chapter Three and corrections for these have been applied to the data to remove any systematic errors. The remaining explanations for the differences observed in the comparisons of F_H and F_E are discussed in the following sections.

4.4.2 Flux Divergence

Previous comparisons between aircraft and ground-based measurements (Desjardins et al., 1989; Desjardins et al., 1992b) have suggested that the divergence of the turbulent fluxes with height as a possible systematic error in aircraft flux measurements and that this could lead to under-estimation of the fluxes compared to ground-based observations. This section examines the effect of flux divergence in the context of the present study.

The low-level flights (6 m) were designed to minimise the effect of flux divergence on the aircraft measurements and divergence is not a plausible explanation for the apparent under-estimation of F_H, and over-estimation of F_E, by the aircraft instruments. For the grid flight comparisons, the aircraft altitude ranged from 15 to 65 m. An estimate of the divergence of F_H with height can be made by assuming a linear decrease in the flux with increasing height throughout the boundary layer to a value of 0 at $0.8z_i$ (Mann and Lenschow, 1994). For F_H equal to 150 W m^{-2} and z_i equal to 1500 m, typical of conditions during the comparison flights (Cleugh et al., 2004), this leads to a decrease of 8% in the flux between the surface and 65 m. This is not consistent with the observation that for these flights the aircraft measurements of F_H are 12% larger than those from the ground-based systems and suggests that flux divergence does not play an important role in the bias observed in these comparisons.

It should be noted that during the work presented here the aircraft altitude never exceeded 65 m. This is lower than the altitudes frequently used by research aircraft measuring fluxes in the boundary layer. Although it is possible to conclude that flux divergence can not explain the systematic error found in the flux comparisons presented here, divergence may still be a plausible explanation when comparisons are attempted with aircraft altitudes of the order of 100 m or greater.

4.4.3 Sampling Limitations

The highest non-dimensional frequency sampled by the ground-based system at the Wattles was $n = fz/u = 5 \times 4.5/3 \approx 8$, assuming a typical wind speed of 3 m s^{-1} and a measurement height of 4.5 m. Similar analysis for the Wagga, Browning and Urana sites gives an upper frequency of $n \approx 15$ for the ground-based data. The lowest frequency sampled by the ground-based systems is determined by the averaging times, 30 minutes for the Wagga and Wattles sites and 15 minutes for the Browning and Urana sites. Using the same typical values as above, this gives the lower bound as $n \approx 0.0008$ and $n \approx 0.0016$ for these sites respectively.

The loss of spectral and cospectral power due to the above sampling restrictions can be estimated using the spectral and cospectral forms published in Kaimal and Finnigan (1994). For the ground-based sites, this yields reductions of about 1% and 2% for variances and covariances respectively. These losses are small compared to other sources of systematic error and no attempt has been made to correct the ground-based observations.

Temperature spectra and $w\theta$ and wq cospectra calculated from the aircraft data (see Chapter Three) suggest that the relatively short averaging lengths used in this study have not caused significant loss of low frequency contributions to the variances and covariances. The water vapour spectra do not approach zero with decreasing frequency, indicating unresolved mesoscale contributions to the water vapour variance, but no correction for this has been attempted here. All spectra and cospectra calculated from the aircraft data show significant loss of high frequency contributions due to under-sampling and phase differences between w and the measured scalars. The aircraft observations have been corrected for this loss, see Chapter Three, but significant bias still remains in the F_H and F_E comparison at the Wattles site and in the F_H comparison at the Browning site. The following section examines surface heterogeneity as a possible reason for the observed bias in these comparisons.

4.4.4 Source-Area of the Measurements

The source-area that contributes a specified fraction of the observed flux can be calculated using approximate analytical solutions to the 2-dimensional advection-diffusion equation (van Ulden, 1978). Source-areas contributing 80% of the flux measured by the aircraft and ground-based instruments have been calculated using the cross-wind concentration distribution from Horst and Weil (1992) and the approximate form for the cross-wind integrated source-area given in Horst and Weil (1994). Full details of the method are given in Chapter Six.

The upwind location of the maximum contribution to the ground-based and aircraft observations, the source-areas contributing 80% of the fluxes and the ratio of the source-areas are listed in Table 4.6 for the low-level, grid and transect flights. An aircraft altitude of 25 m has been used for the grid and transect flights. Typical values of $L = -15$ m and $\sigma_{WD} = 30$ °, where σ_{WD} is the standard deviation of the wind direction, have been used to calculate the source-areas based on observations during the comparison periods. For these conditions, an aircraft at 25 m will be in the convective matching layer (Kaimal and Finnigan, 1994). The assumptions used to derive the source-area weight function may no longer be valid in this layer and the results for this aircraft height are indicative only. Source-areas for the aircraft measurements are calculated by multiplying the upwind dimensions of the source-area by the length of the aircraft run. These were 1230 m, 59 km (sum of all grid leg lengths) and 10 km (the averaging length) for the low-level, grid and transect flights respectively. The source-area for a ground-based site is an ellipsoid enclosing the smallest surface patch that contributes 80% of the flux and this is multiplied by the number of sites averaged to get the total area of the ground-based observation.

The source area of the aircraft observations during the low-level comparisons was typically 30 times larger than that for the ground-based observations, even though the measurement heights were similar. This is because the source area for the aircraft measurements is a broad swathe that extends the length of the aircraft run, in contrast to the ground-based footprint whose width is determined by the lateral turbulence intensity, see Kaharabata et al (1997). The difference between the source areas for the aircraft and the ground-based measurements is even more pronounced for the grid and transect comparisons where the source area of the aircraft observations was 1250 and 420 times larger than the total for the ground-based sites. The larger surface areas influencing the aircraft observations mean that surface heterogeneity not sampled by the ground-based instruments may introduce bias into comparisons of aircraft and tower data. The effect of this is examined in the next section.

Table 4.6 Upwind locations of the maximum contribution and source-areas contributing 80% of the flux for the low-level, grid and transect flights. z_0 is the roughness length, $X_{m,GB}$ and $X_{m,AC}$ are the upwind locations of the maximum contribution and A_{GB} and A_{AC} are the source-areas for the ground-based and aircraft observations respectively. A_{AC}/A_{GB} is the ratio of the source-areas for the ground-based and aircraft observations.

	z_0 (m)	$X_{m,GB}$ (m)	$X_{m,AC}$ (m)	A_{GB} (m^2)	A_{AC} (m^2)	$\frac{A_{AC}}{A_{GB}}$
Low-level	0.12	11	15	4.6×10^3	1.4×10^5	30
Grid	0.06	13	131	2.8×10^4	3.5×10^7	1250
Transect	0.06	13	131	1.4×10^4	5.9×10^6	420

4.4.5 The Effect of Surface Heterogeneity

To check the effect of surface heterogeneity, F_H and F_E were calculated from a subset of data when the aircraft was immediately adjacent to the towers and compared to the fluxes for the whole low-level pass, grid or transect segment. Aircraft data at the Wattles site were subdivided into times when the aircraft was over the instrumented field and when it was over the adjoining, non-instrumented field and the fluxes calculated for each data subset. The flux for the whole run was calculated as the average of the fluxes over the instrumented and non-instrumented fields to avoid biasing the comparison by including longer wavelength contributions in the whole run fluxes.

For the grid flights, the aircraft data were divided into 2 km non-overlapping blocks and the fluxes for the instrumented sites were calculated from the blocks centred on the ground-based sites. Fluxes for the whole grid were calculated by averaging the non-overlapping 2 km blocks to avoid biasing the comparison by inclusion of long wavelength contributions in the whole grid fluxes. For the transect flights, the data from each 10 km segment used in the comparison of the ground-based and aircraft observations was split into 2 km blocks. The fluxes for the 2 km block closest to the ground-based sites were then compared to the average of the five blocks comprising the 10 km segment. The use of such short averaging lengths will result in the loss of approximately 10% to 15% of the turbulent fluxes but this is tolerated here because the emphasis is on identifying spatial patterns in the fluxes and not on the absolute accuracy of the fluxes. The sensitivity of this method to the choice of averaging length was checked by varying the averaging length from 1500 to 3000 m in 500 m increments but no significant change in the results was observed.

The effect of surface heterogeneity on the comparison was assessed by calculating the ratio of the flux over the instrumented site, F_S^I, to the flux over the whole low-level pass, grid pattern or transect segment, F_S, where S can be either H or E. A ratio greater than one indicates that the flux measured over the instrumented site is greater than the flux measured over the whole aircraft flight track used in the comparison and means that the aircraft data will underestimate the ground-based observations by the inverse of this ratio. The converse is true for a ratio of less than one and a ratio of unity implies there is no systematic difference in the data due to surface heterogeneity. The aircraft observations were multiplied by the ratio F_S^I / F_S to correct for the effect of surface heterogeneity and the corrected fluxes compared once more to the ground-based values. The results of this analysis are presented in Table 4.7 for the low-level, grid and transect flights.

There are significant differences between the aircraft measurements of F_H and F_E over the instrumented sites and the fluxes measured over the whole low-level pass and a similar result is found for F_H in the grid flight comparison. There is no evidence of significant bias due to surface heterogeneity in the comparison of F_E for the grid flights or in the comparison of either flux for the transect flights. Correction of the aircraft observations of F_H and F_E from the low-level flights eliminates virtually all of the observed bias with the mean difference between the aircraft and ground-based values being reduced from 24 W m^{-2} to 7 W m^{-2} and from -16 W m^{-2} to

-7 W m^{-2} for F_H and F_E respectively. This is also true for the comparison of F_H from the grid flights where the mean difference is reduced from -22 W m^{-2} to -6 W m^{-2}. The slope of the line of best fit through the origin for these comparisons is no longer significantly different from unity.

Table 4.7 Results of the surface heterogeneity analysis described in the text and statistics for the comparison of F_H and F_E after correcting for the effect of the observed heterogeneity. F_S^I / F_S is the ratio of the flux over the instrumented site (F_S^I where S can be either H or E) to the flux over the whole low-level pass, grid pattern or transect segment used in the comparison (F_S). Other columns are as for Table 4.2.

		F_S^I / F_S	m	r^2	MD W m^{-2}	SD %
Low-level	F_H	1.11	0.97 ± 0.08	0.90	7	20
	F_E	0.89	1.02 ± 0.07	0.82	-7	15
Grid	F_H	0.91	1.01 ± 0.07	0.74	-6	18
	F_E	1.03	1.05 ± 0.04	0.89	-12	15
Transect	F_H	0.98	1.07 ± 0.07	0.80	-27	34
	F_E	1.02	0.95 ± 0.07	0.66	5	31

The results show that most of the bias observed in the comparison of aircraft and ground-based observations of F_H and F_E can be attributed to the effect of surface heterogeneity and the differing source-areas of the measurements. Excellent agreement between aircraft and ground-based observations is obtained when corrections for the surface heterogeneity are applied to the aircraft data.

4.5 Summary and Conclusions

Measurements of means, standard deviations and fluxes of various micrometeorological quantities at four locations were compared with aircraft measurements made during the 1995 OASIS field campaign. Aircraft measurements were made at 6 m over adjacent fields at the Wattles site. The low level of the aircraft flights allowed direct comparison of means and standard deviations measured by the aircraft and the ground-based systems in addition to the comparison of the fluxes of sensible heat, latent heat, net radiation and the friction velocity. The ground-based site at Browning consisted of four systems located in adjacent crop and pasture fields over which the aircraft flew grid patterns at between 15 and 65 m. Data collected by the aircraft in the vicinity of the Wagga, Browning, and Urana sites during the transect flights at 20 m were also used to compare aircraft and ground-based measurements. The height difference between the aircraft and ground-based measurements during the grid and transect flights restricts the comparison to the fluxes only.

Aircraft and ground-based observations of the mean wind speed, wind direction and air temperature agreed to within the uncertainties in the calibration of aircraft and ground-based instruments. Observations of σ_w, σ_u, σ_θ and σ_q also agree after correcting the aircraft temperature measurements for inadequate sensor response time and correcting all variances for inadequate sampling frequency. The scatter in the standard deviation comparison is similar to that observed between ground-based instruments during ITCE76. Approximately half of the scatter can be attributed to the random error in the aircraft and ground-based measurements.

Excellent agreement is obtained between the aircraft and ground-based measurements of F_H and F_E when the aircraft data are corrected for the attenuation of high frequency contributions and the combined effect of differing source-areas and surface heterogeneity. The attenuation of high frequency contributions arises from inadequate temperature sensor response time and an increasing phase difference between the aircraft w and all scalar measurements at frequencies above 3 Hz. When the effects of surface heterogeneity are included, the mean difference between the aircraft and ground-based measurements of F_H for the low-level, grid and transect flights are 7, -6, -27 W m^{-2} respectively and the mean differences in F_E are -7, -12 and 5 W m^{-2} respectively. When combined with a simple model of F_G, the aircraft measurements of F_H and F_E perform as well as the ground-based data in closing the surface energy budget in terms of both bias and scatter.

Aircraft measurements of the friction velocity agreed well with the ground-based values when the aircraft altitude satisfied $|z/L| < 1$ and the aircraft was within the surface layer. This condition was satisfied for most of the low-level flights. Agreement was poor when the aircraft altitude exceeded the $|z/L| < 1$ condition and the aircraft was within the free convection layer, as was the case for most of the grid and transect flights. The implications of this are discussed in Chapter Five.

The comparison of $S_{\downarrow}$ showed that the aircraft instruments overestimated this quantity by 10% with respect to the ground-based systems. The reason for this discrepancy has not been explored but it is likely to be due to differences in the calibration procedures for both sets of instruments. All aircraft measurements of $S_{\downarrow}$ have been reduced by 10% to ensure the compatibility of these data in subsequent analyses.

Comparison of F_C shows that the aircraft sensors under-estimate this quantity with respect to the ground-based systems, even after correction for inadequate sensor response time. This is most likely due to surface heterogeneity and the differing source-areas influencing the aircraft and ground-based observations. The comparison is dominated by a small number of data points recorded at the Wagga site in conditions that would tend to exacerbate differences in the photosynthetic activity of the surfaces sampled by the aircraft and the ground-based measurements. Evidence for significant differences in the surfaces sampled by the Wagga crop and pasture systems and the aircraft is presented in Chapter Six and Chapter Seven. The aircraft measurements of F_C are used without further correction in subsequent analyses.

This comparison shows that aircraft observations of means, variances and fluxes agree well with ground-based measurements and that the aircraft data performs as well as the ground-based data in closing the surface energy budget. This allows a high level of confidence to be placed in the aircraft data. Scatter in the comparisons is generally larger than expected but similar in magnitude to that observed during ITCE76. Aircraft measurements of the turbulent fluxes at heights up to 65 m are shown to be influenced by surface heterogeneity and correction for this removes most of the bias in the comparisons of F_H and F_E. Inadequate sensor response times, especially for temperature and CO_2 sensors, and ground-based sites that are not representative of their surroundings need to be considered as plausible explanations for the bias observed in previous comparisons of aircraft and ground-based measurements. The results presented in this chapter demonstrate that with careful attention to sensor performance, correction of losses due to inadequate sampling and allowance for the effect of surface heterogeneity there are no large, systematic errors in aircraft observations. The conclusion is that aircraft and ground-based observations are equivalent and that data from the two sources can be integrated with confidence.