

## Choosing the Averaging Interval for Comparison of Observed and Modeled Fluxes along Aircraft Transects over a Heterogeneous Surface

MARGARET A. LEMONE

*National Center for Atmospheric Research,\* Boulder, Colorado*

ROBERT L. GROSSMAN

*PAOS, University of Colorado, Boulder, Colorado*

FEI CHEN AND KYOKO IKEDA

*National Center for Atmospheric Research,\* Boulder, Colorado*

DAVID YATES

*National Center for Atmospheric Research,\* and University of Colorado, Boulder, Colorado*

(Manuscript received 22 February 2002, in final form 28 August 2002)

### ABSTRACT

Data from the April–May 1997 Cooperative Atmosphere Surface Exchange Study (CASES-97) are used to illustrate a holistic way to select an averaging interval for comparing horizontal variations in sensible heat ( $H$ ) and latent heat (LE) fluxes from low-level aircraft flights to those from land surface models (LSMs). The ideal filter can be defined by considering the degree to which filtered aircraft fluxes 1) replicate the observed pattern followed by  $H$  and LE at the surface, 2) are statically robust, and 3) retain the heterogeneity to be modeled. Spatial variability and temporal variability are computed for different filtering wavelengths to assess spatial variability sacrificed by filtering and how much temporal variability can be eliminated; ideally, spatial variability should approach or exceed temporal variability. The surface pattern to be replicated is a negative slope when  $H$  is plotted against LE for a given time. This is required for surface energy balance if  $H$  or LE vary horizontally more than their sum,  $R_n - G$ , the difference between the net radiation and heat flux into the ground. Statistical confidence is estimated using conventional techniques. The same factors can be used to examine comparisons already done, or to estimate the number of flight legs needed to measure heterogeneity at a given scale in future field programs.

### 1. Introduction

Using data from the 1997 Cooperative Atmosphere Surface Exchange Study (CASES-97), Chen et al. (2003) compare sensible and latent heat fluxes ( $H$  and LE) measured along low-level flight legs to surface fluxes along the flight tracks based on three land surface models (LSMs). This paper is motivated by two of their results: 1) while  $H$  and LE averaged over the entire flight tracks were consistent with LSM predictions allowing for possible low biases in aircraft  $H$  (LeMone et al.

2002),<sup>1</sup> the LSMs were not as successful at reproducing the *variability* along the flight tracks; and 2)  $H$  and LE estimated along aircraft low-level legs were positively correlated spatially, such that their sum locally exceeded (or was well below) what would be expected from a local surface-energy balance. Here we explore the reasons for these two findings and suggest means to improve strategies for future field experiments that maximize the suitability of aircraft flux estimates for comparisons to modeled or observed horizontal variability at the surface.

Given the large sample needed to obtain aircraft-derived fluxes to a desired accuracy, the relatively greater success of Chen et al. (2003) in model-aircraft flux com-

---

\* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

---

Corresponding author address: Dr. Margaret A. LeMone, NCAR, P.O. Box 3000, Boulder, CO 80307.  
E-mail: lemone@ucar.edu

---

<sup>1</sup> Low biases in aircraft  $H$  estimates are commonly reported [e.g., low values of  $H$  in the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE; Sellers et al. 1992) by Kelly et al. (1992)].

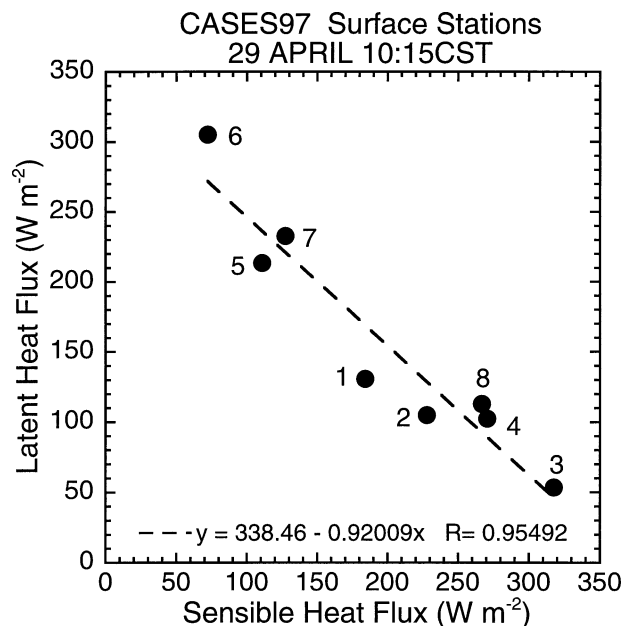


FIG. 1. For 29 Apr,  $H$  vs  $LE$  from eight NCAR surface-flux stations. Surface cover: dormant grass for sites 1, 2, and 8; rapidly growing winter wheat for sites 5, 6, and 7; and bare ground for sites 3 (corn and soybean seedlings by 20 May) and 4 (scattered weeds by 20 May).

parisons for whole flight legs as opposed to variations along the flight track is not surprising. Mahrt (1998) and others have noted that the number of aircraft passes typically flown to sample average fluxes is less than that needed to represent horizontal variability in surface fluxes reliably. From Eq. (17) in Mann and Lenschow (1994), the uncertainty in flux  $F$ ,  $\sigma_F \propto L_x^{-1/2}$ , where  $L_x$  is the length of the record, so that the uncertainty for a flux averaged over a fourth of the flight leg will be twice that for the whole leg, and so on. This makes it more difficult to obtain an aircraft flux estimate of sufficient accuracy for comparison to surface-flux variability at smaller scales. Scatter in the flux data is sometimes reduced by high-pass filtering the data before fluxes are calculated. For example, in FIFE many investigators (e.g., Desjardins et al. 1992) high-pass filtered the data (cutoff 3–5 km) before calculating the fluxes. This procedure retains smaller eddies presumably more closely tied to the local landscape but could miss some flux contribution from the larger eddies.

The second finding of Chen et al. (2003) motivates a closer look at how the aircraft data should behave for meaningful comparisons to surface observations or model results. At the surface,

$$H + LE \approx R_n - G, \quad (1)$$

where  $H$  is sensible heat flux,  $LE$  is latent heat flux,  $R_n$  is net radiation, and  $G$  is the heat flux into the ground. If the spatial variation in  $R_n - G$  is less than that for  $H$  or  $LE$ , the latter two quantities should be negatively correlated, as is illustrated in Fig. 1. The negative  $\Delta LE/\Delta H$

slope in the figure is not surprising, given nearly clear skies and the association of high  $LE$  and low  $H$  with rapidly growing unstressed winter wheat (sites 5–7), lower  $LE$  and higher  $H$  with mostly dormant grass (sites 1, 2, 8), and the highest  $H$  and lowest  $LE$  with dry, bare ground (sites 3, 4). It follows that smoothed  $H$  and  $LE$  from an aircraft flying along a track 30 m above similar surface vegetation should exhibit a similar negative slope. Since (1) requires adherence to the surface energy budget, we call this test the surface-energy budget constraint.

Surface fluxes of  $H$  and  $LE$  from FIFE show a similar relationship. Based on three years of growing season data, Fritschen and Qian (1992) tabulate the accumulated components of the surface energy budget for six stations scattered around the domain. For all three years,  $H$  and  $LE$  are negatively correlated, with  $\Delta LE/\Delta H$  slopes of  $-2.5$  ( $R = 0.88$ ) for 26 May–16 August 1987,  $-1.4$  ( $R = 0.89$ ) for 10 May–18 September 1988, and  $-1.1$  ( $R = 0.87$ ) for 21 July–13 August 1989. The slopes of the relationship are related to precipitation: 1989 was the wettest period, and 1987, including only a few weeks in midsummer, was the driest. (Figure 1, which has slope most like the driest year, is based on data collected over a week after any precipitation.)

We show here how adherence to (1) can be used along with statistical uncertainty and comparisons of variability in space and time to select spatial averaging lengths for comparing aircraft fluxes to those predicted from LSMs. Alternatively, these three factors can be considered in laying out the flight tracks and estimating the number of flight legs needed in future field programs for good comparisons.

## 2. Data and data processing

These data were collected as part of the CASES-97 field program, in the lower Walnut River watershed east of Wichita, Kansas, between 21 April and 21 May (LeMone et al. 2000). The experimental array is shown in Fig. 2. The surface is covered primarily by grassland in the eastern part of the watershed, with crops (primarily winter wheat, some corn, and soybeans) and human settlement more common in the western part. We focus on 29 April, 10 May, and 20 May, whose characteristics are summarized in Table 1. These days marked the transition from highly heterogeneous vegetation (dormant grass, rapidly growing winter wheat, bare soybean and corn fields) to more homogeneous (grass and winter wheat both green, sparse corn and soybean seedlings). Further, there were varying degrees of dry down, with over a week without rain on 29 April, two days without rain on 10 May, and heavy rain the day before 20 May, which lead to standing water in the fields. Mean boundary layer winds varied from 5–6 m s<sup>-1</sup> on 10 May, to 10 m s<sup>-2</sup> on 29 April. Weather was sunny, with clear skies to broken cirrus clouds.

The aircraft data are the same as used in Chen et al. (2003), namely, fluxes from the Wyoming King Air and

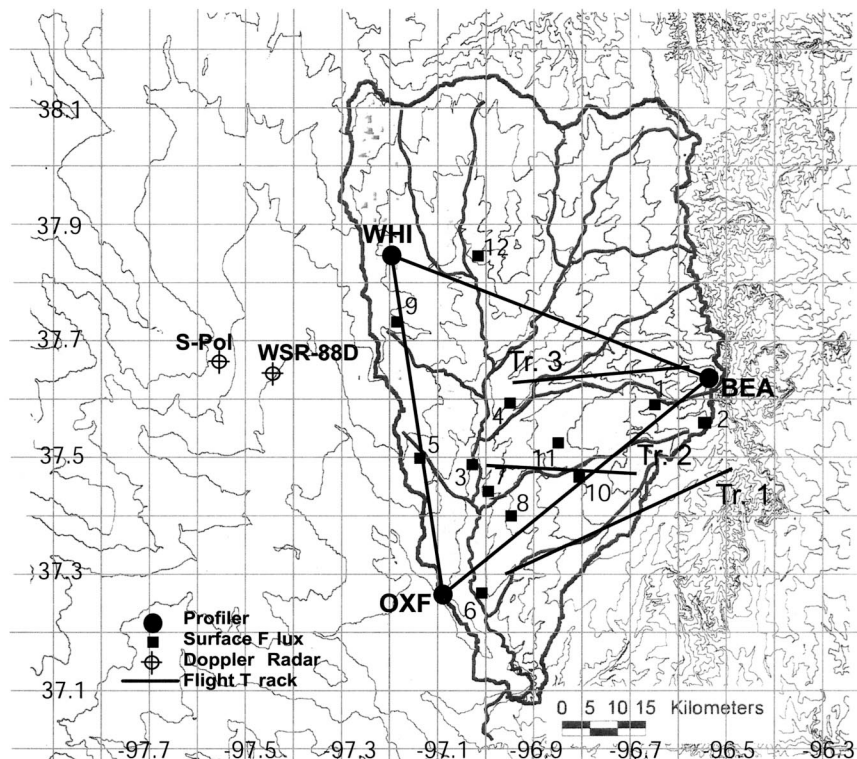


FIG. 2. Flight tracks, superposed on the CASES-97 instrument array. Contours are at 20-m intervals. For reference, Beaumont (BEA) is at 478 m, Whitewater (WHI) is at 420 m, and Oxford (OXF) is at 360 m. Numbers denote locations of surface stations; stations 1–8 are represented in Fig. 1. The interval  $0.01^\circ$  longitude corresponds to 0.97 km for track 1, 0.88 km for track 2, and 0.89 km for track 3.

the National Oceanic and Atmospheric Administration (NOAA) Twin Otter, collected on repeated straight-and-level flight tracks (Table 2; Fig. 2),  $\sim 30$ – $60$  m above the surface. For the King Air, these legs are the lowest of four straight-and-level legs flown over the same track at alternating headings in descending “stacks,” starting just below the top of the mixed layer. The Twin Otter alternated heights between just below the top of the mixed layer and  $<60$  m above the surface. Both aircraft performed soundings to find the mixed-layer top. Flight patterns typically lasted about 4 h, from 0830 to 1230 LST. The King Air typically flew four low-level legs (three on 20 May), and the Twin Otter flew four to six. Each aircraft flew along its assigned track to within  $\sim 1$  km in the north–south direction. Average near-surface estimates of LE from the two aircraft are similar, but near-surface estimates of  $H$  from the Twin Otter are 30% higher than those for the King Air (LeMone et al. 2002).

The three flight tracks extended over a mix of crops, trees, and grass. Grassland dominated the eastern half to two-thirds of all three tracks, while a mix of crops (25–50%) and grass (50–75%) lay beneath the western portion, following the general pattern in the watershed. Each track extended over groups of trees, concentrated

in riparian zones or along section- or quarter-section boundaries. None of the flight tracks extended over urban areas. Further detail on land use can be found in Chen et al. (2003).

It was assumed in the Chen et al. (2003) study that the model fluxes for the area directly beneath the flight track could be compared with the aircraft fluxes. Based on Horst and Weil (1992) and Desjardins et al. (1992), estimates of the upstream distance for which surface and flight-level fluxes should be most highly correlated were less than 1 km. Chen et al. (F. Chen 2002, personal communication) tried to account for upstream fetch by comparing aircraft data to modeled fluxes 1 km upstream from the flight track, with little difference in results from those presented. Perhaps this is because the land-use gradient is similar across the watershed. However, J. Song (2002, personal communication) found that accounting for the flux footprint improves model-aircraft intercomparisons using the same dataset.

The surface energy fluxes are from eight Portable Automated Mesonet (PAM) III eddy-correlation surface-flux stations operated by the National Center for Atmospheric Research (NCAR). These stations, numbered 1–8 in Fig. 2, were sited in rough proportion to

TABLE 1. Characteristics of the three days compared.

Date	Wind (m s <sup>-1</sup> )	Cloud	Time since last rain	IR sfc temp* (°C)	Grass	Winter wheat (height in cm)	$\Delta LE/\Delta H$ sfc stations
28–29 Apr	S 10**	Broken cirrus**	> 7 days	35 ± 4.9**	Dormant	30–45 (1 May)	-0.9 ± 0.15
10–11 May	SSW 5-6	Clear	2 days	33 ± 4.6	Some green	60–90 (12 May)	-1.1 ± 0.49
20–21 May	ENE 6-7	Scattered cirrus	1 day	29 ± 2.7	Green	60–90, maturing (20 May)	+0.5 ± 0.27

\* Uncorrected Heiman downward-looking radiometer data from low-level King Air flight legs.

\*\* Conditions are for 29 Apr.

land-use characteristics (grass, winter wheat, bare ground; Fig. 1) and over a range of elevations. At the beginning of the period (Table 1), the grass was nearly dormant, the winter wheat was growing actively and bare ground was truly bare. However, by the end of the period, the winter wheat was starting to mature, the grass was green, and the “bare ground” sites had sparse vegetation cover (corn and soybean seedlings at one site; weeds at the other). Further detail on siting, instrumentation, and data processing can be found in LeMone et al. (2000), Horst and Oncley (1995), and Millitzer et al. (1995).

The analysis in this paper is based on 0.01° longitude (about 1 km; see Fig. 2) block averages of the flux of  $s$  = potential temperature  $\theta$  or mixing ratio  $q$  along each low-level flight leg. These averages were found by (a) linearly detrending  $s$  and  $w$  over the flight leg, such that the detrended quantities  $s'$  and  $w'$  average to zero over the leg, (b) forming the product  $w's'$ , (c) block averaging  $w's'$  over 0.01° longitude intervals, and (d) converting the 0.01° block averages of  $w'\theta'$  to sensible heat flux  $H$  and  $w'q'$  to latent heat flux LE, using 0.01°-longitude block-averaged values of air density and the appropriate constants. Because the entire leg has been used in detrending, the average of the fluxes evaluated for the individual 0.01° segments is equal to the leg average flux.

All the low-level tracks in each pattern are also time averaged at common longitude points to form a grand average leg for comparison to surface fluxes.

### 3. Analysis techniques and results

Our objective is to select the right smoothing to produce a horizontal transect of  $H$  and LE from low-altitude

aircraft data that is optimum for comparison with observed and modeled surface values. Such a transect should provide a significant range of flux values that are physically reasonable and statistically robust. All of these attributes can be designed into a field program, but after a field program the only options are to average sufficiently in space and time that the uncertainty is acceptable and to choose flight data that meet the other criteria after adequate averaging. The amount of temporal averaging is fixed by the number of flight legs, but we can change the spatial filter. In this section, we describe three techniques that provide useful information for selecting the proper spatial filter: (a) looking for the negative correlation between  $H$  and LE required for surface-energy balance in heterogeneous conditions, (b) estimating statistical uncertainty, and (c) comparing spatial variability to temporal variability, as a function of filter wavelength.

Spatial averaging is done with a running-mean filter for simplicity and to minimize the number of weights needed, even though the response function oscillates rather than making a smooth transition from 1 to 0 between the cutoff and terminating wavelengths. As illustrated in Fig. 3, the 0.01° averaged fluxes have considerable horizontal variability, which reflects local concentration of fluxes in individual events as well as the fluxes at the surface beneath the aircraft. In the figure, running averages of 2, 3, and 4 points are compared to the original curve. As expected, the smaller-scale variations are filtered out. This sacrifices surface-correlated flux variability at small scales but potentially isolates variability at larger scales with acceptable uncertainty. Since the entire flight track is used in detrending the data, the flux for the smoothed data will approach the leg average flux.

TABLE 2. CASES-97 aircraft data used in analysis.

Date	King Air grand average legs					Twin Otter grand average legs				
	Track	Legs	$n_{\theta}^{\#}$	Lat (°N)/lon (°W)		Track	Legs	$n_{\theta}^{\#}$	Lat (°N)/lon (°W)	
				West end	East end				West end	East end
29 Apr	1	4	46	36.306/96.960	37.473/96.510	3	6	30	37.640/96.870	37.665/96.580
10 May	1	4	42	37.312/96.930	37.468/96.520	3	4	19	37.640/96.860	37.658/96.680
20 May	3	3	38	37.620/96.940	37.643/96.570	2	5	16	37.487/96.985	37.475/96.865

\* Total number of 0.01°-longitude leg segments; the number of nonoverlapping samples for a four-point running average would thus be the number of segments divided by 4.

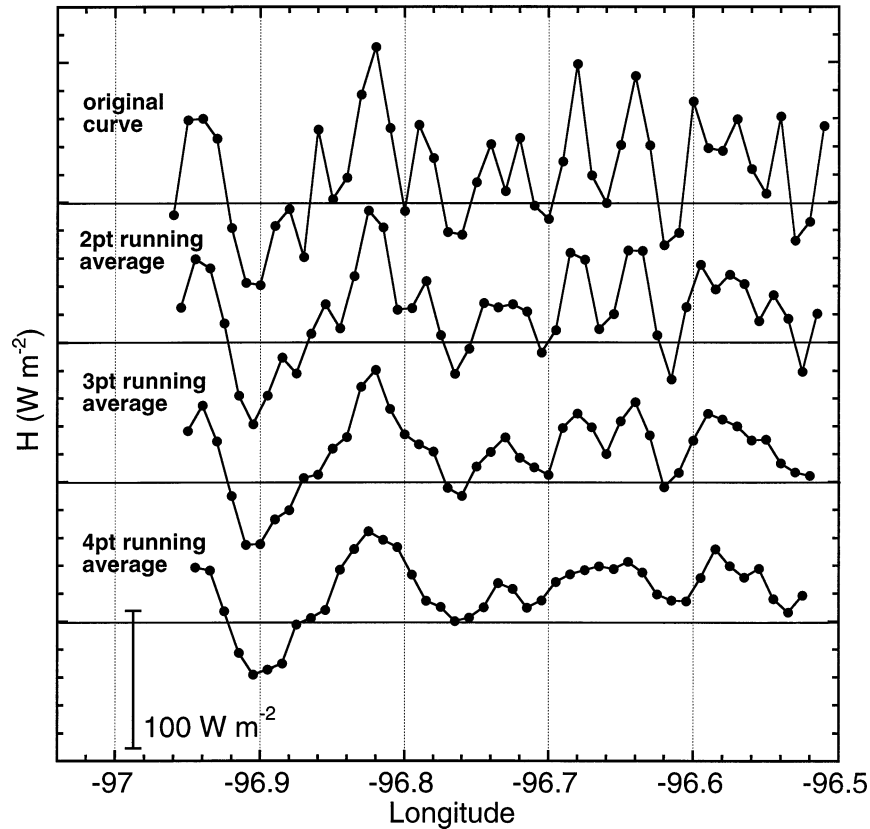


FIG. 3. For 29 Apr, comparison of fluxes for  $0.01^\circ$  average, and running averages taking 2, 3, and 4 points. The curves are separated by  $100 \text{ W m}^{-2}$ . One point corresponds to  $\sim 1 \text{ km}$ .

We look for replication of the observed surface  $\Delta H/\Delta \text{LE}$  slope, evaluate statistical uncertainty, and compare temporal to spatial variability using the original  $0.01^\circ$  longitude data, to the two-point ( $0.02^\circ$  longitude) running averages, and so on, until the data were smoothed with running averages extending over a half the leg. As illustrated in Fig. 3, our filtering method shortens the smoothed leg more as larger numbers of points are included in the running means. We use the terms “filter length,” “filter wavelength,” or “filter interval” in referring to the number of points in the running mean, where each point represents  $0.01^\circ$  longitude or roughly 1 km for the track headings in Fig. 2.

#### a. Energy balance constraint

For this constraint to be useful, there must be sufficient horizontal variability in the surface data to produce a well-defined negative slope  $\Delta \text{LE}/\Delta H$ . If the stations upon which the  $H$ – $\text{LE}$  relationship is based do not lie within the footprint for the flight track, then it is necessary to confirm that the flight track samples the important land-use categories. Provided these criteria are satisfied, the model fluxes beneath the flight track should follow the same relationship as the observations, as should the aircraft fluxes at 30 m—provided there is an

adequate statistical sample. The relationship will be affected if  $H$  and  $\text{LE}$  decrease with height at different rates, but this effect at 30 m should be small.

Figure 4 compares the slopes of the least squares best-fit straight lines relating  $H$  and  $\text{LE}$  for the eight PAM III surface-flux stations at 1115 CST for 29 April, 10 May, and 20 May. Two of the three have the  $\sim -1$  slope of Fig. 1, while the slope on 20 May is positive but indeterminate because of the small spread of the points.<sup>2</sup> The best fit, on 29 April, is related to large horizontal variability in  $H$  and  $\text{LE}$ . From Table 1 and Fig. 1, this variability relates to the greatest surface heterogeneity of the three days: grasses are still mostly dormant, winter wheat is growing rapidly, and no rain has fallen for over a week. In contrast, the vegetation on 20 May is almost uniformly green and near-surface soil moisture nearly saturated due to heavy rains the day before.

Figure 5 shows that the modeled fluxes beneath the flight track using the three LSMs of Chen et al. (2003) show some similar features to the surface station fluxes. This is to be expected, since the flight tracks extend

<sup>2</sup> According to Yates et al. (2001), the imbalance in the surface-energy budget in the CASES-97 surface dataset reached  $100 \text{ W m}^{-2}$ , implying that the horizontal variability in the fluxes on 20 May could be related to measurement uncertainty rather than real variability.

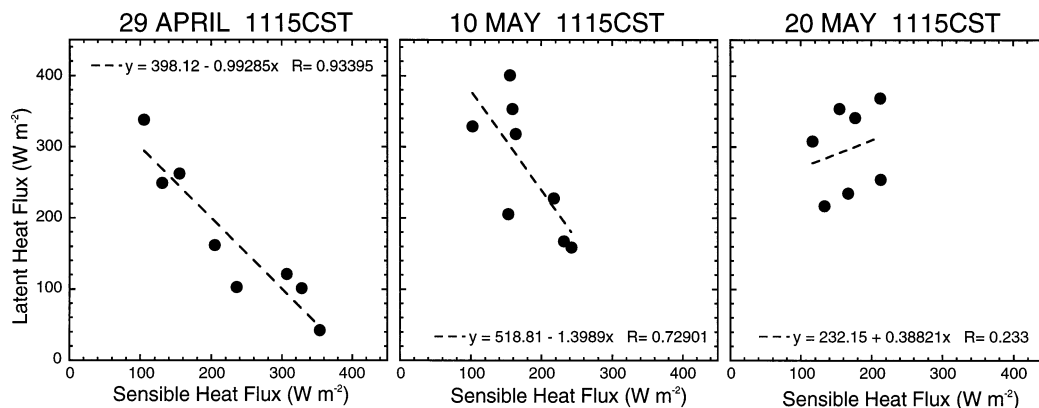


FIG. 4. Comparison of surface-station-based best-fit lines for LE as a function of  $H$ , for 29 Apr, 10 May, and 20 May for the half-hour period centered at 1115 CST. For surface conditions, see Table 1.

over a similar range of land-use categories. All models on all days show negative  $\Delta LE/\Delta H$  slopes. Consistent with the surface observations, the horizontal range of modeled  $H$  and LE is greatest on 29 April, and least on 20 May. The  $\Delta LE/\Delta H$  slope tends to be closer to  $-1$  on 29 April and 10 May, as expected.

To see whether the variation in  $R_n - G$  is less than that of  $H$  or LE, we examined both observations and model results. The “observed” values of  $R_n - G$ , based on  $R_n$  from aircraft and  $G$  averaged from the three LSMs, varied  $<50$ – $60 W m^{-2}$  on all three days, consistent with the lack of thick clouds (Table 1). This range is smaller than the range of  $H$  and LE at the surface stations and for the NCAR LSM for both 29 April and 10 May (track 1). Table 3 compares the LSM standard deviations of  $H$ , LE, and  $H + LE = R_n - G$  along the *King Air* tracks. The NCAR LSM most closely matches expectations based on the surface data, with  $\sigma_{H+LE} < (\sigma_H, \sigma_{LE})$ , and  $\Delta H/\Delta LE$  closest to  $-1$  for 29 April and 10 May. For the two days for which negative correlation is expected, only the Oregon State University (OSU) LSM on 10 May has  $\sigma_{H+LE}$  greater than both  $\sigma_H$  and  $\sigma_{LE}$ , consistent with the poorly defined (albeit negative)  $\Delta LE/\Delta H$ -slope. All three LSMs produce the least horizontal variability in  $H$  and  $H + LE$  for 20 May, as expected; but variability in LE is between values for 29 April and 10 May except for the NCAR LSM.

We have established that  $\Delta LE/\Delta H$  should be negative along the flight tracks for 29 April and 10 May. Thus it is reasonable to assume that aircraft values should follow a similar pattern. However, this is not the case. The plots of  $H$  versus LE for the four low-level *King Air* legs on 29 April in Fig. 6 show considerable scatter, with the only distinct slope a *positive* slope for the leg beginning at 1212:03 CST. Canadian Twin Otter  $H$  and LE data at 90 m from FIFE, in Figs. 6–11 of Desjardins et al. (1992), show similar behavior, with positive or null spatial correlations between  $H$  and LE in the majority of cases. Fluxes were estimated at 3.8-km inter-

vals in a grid over the FIFE domain after high-pass filtering (cutoff frequency 4.5 km).

The lack of relationship between the surface fluxes and those sensed by the aircraft is just what we would expect if the surface effect ( $H$  and LE negatively correlated) is masked by the effect of the atmosphere, which produces a positive correlation between  $H$  and LE. Chen et al. (2003) show an example of an individual event for which  $H + LE + G > R_n$ , probably because converging winds concentrate warm, moist air and then carry it upward. Atmospheric processes are isolated in the plot  $H$  against LE for 20 May in Fig. 7, because of the relatively uniform surface conditions. Rather than showing little relationship, the plots show the positive  $\Delta LE/\Delta H$  slope that reflects the concentration of fluxes by large eddies. In the plots,  $H + LE$  exceeds  $700 W m^{-2}$ , far greater than the modeled (Table 3) and observed values (not shown) of  $R_n - G \sim 500 W m^{-2}$ .

Several researchers have demonstrated that large eddies (typically hundreds of meters to several kilometers across and extending through the boundary layer) tend to concentrate turbulent elements and, hence, fluxes such as  $H$  and LE into their updraft regions. This conclusion is based on observations over water (LeMone 1976; LeMone and Pennell 1976; Khalsa and Businger 1977)—where there was little horizontal nonuniformity on the 10–50-km scale—as well as over land (e.g., Kaimal et al. 1976). The concentration of turbulence/flux into updraft regions is reflected in the correlation of vertical velocity  $w$  with variances of vertical velocity, potential temperature  $\theta$ , or mixing ratio  $q$  (e.g., Lenschow et al. 1980; Young 1988). If  $\theta$  and  $q$  are correlated with  $w$ , positive  $w$  skewness also contributes to greater  $H$  and LE flux in updrafts than in downdrafts. This tendency is reinforced when  $\theta$  and  $q$  are also positively skewed. Positive  $w$  skewness is observed over a wide range of convective boundary layer regimes (e.g., Lenschow et al. 1980; Hunt et al. 1988; LeMone 1990), as is positive  $\theta$  skewness (e.g., Lambert et al. 1999; Mahrt

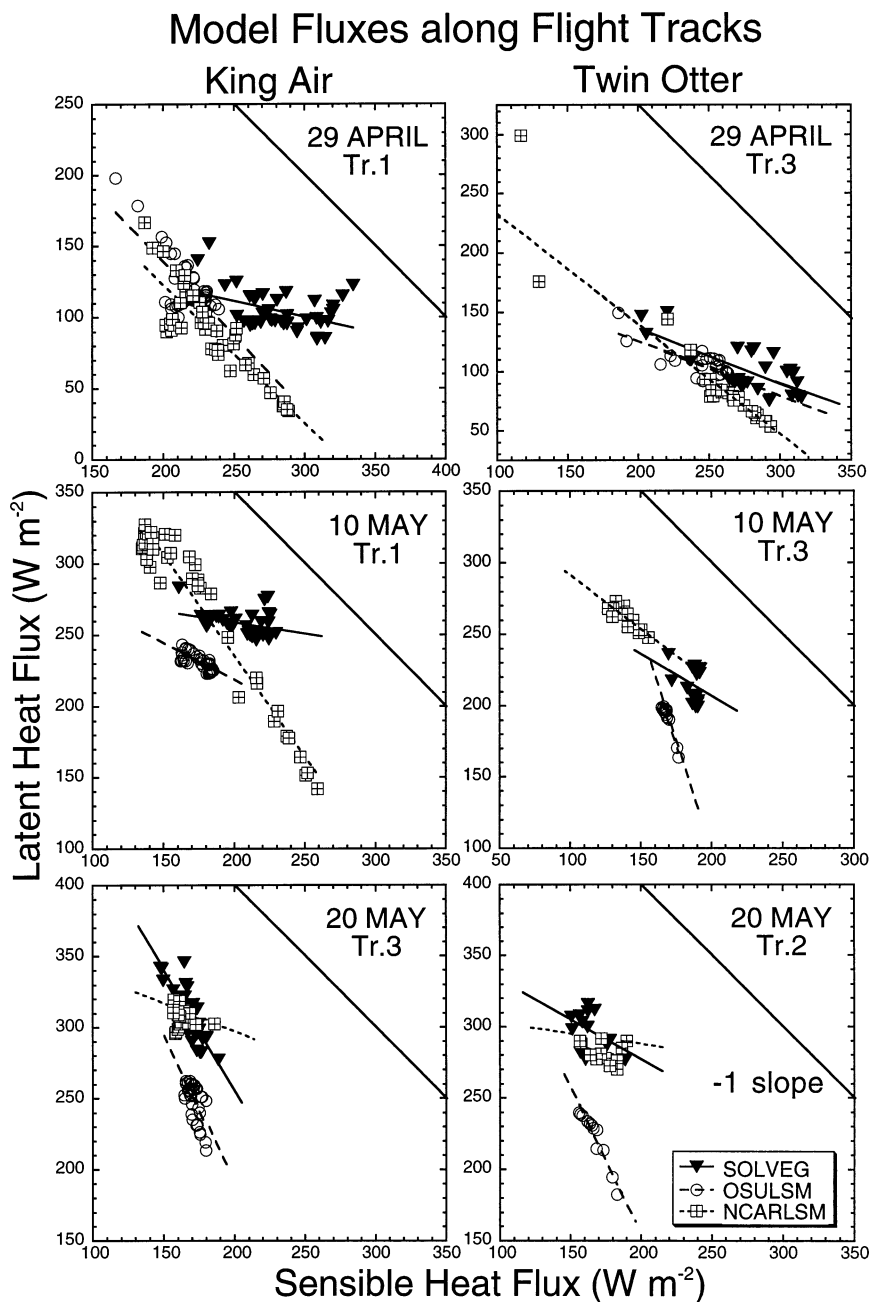


FIG. 5. For flight tracks flown on 29 Apr, 10 May, and 20 May,  $H$  vs  $LE$  for Twin Otter and King Air, based on three LSMs. Model results have been smoothed using a filter designed to have a similar response to the aircraft four-point running mean. The  $-1$   $H$ - $LE$  slope, consistent with surface-energy balance with small horizontal variation in  $R_n - G$ , is plotted for reference. Brief descriptions of the three LSMs are in Chen et al. (2003). For further detail, consult Nagai (2001) for SOLVEG, Bonan (1996) for the NCAR LSM, and Chen et al. (1996) and Pan and Mahrt (1987) for the OSU LSM.

1991) and positive  $q$  skewness at heights not strongly impacted by entrainment (Mahrt 1991; Lambert et al. 1999). Large eddy simulations also show concentration of fluxes in updrafts (e.g., Deardorff 1974; Moeng and Sullivan 1994).

The effect of flux concentration by individual at-

mospheric events is reduced by averaging in time, as is illustrated by less positive slopes for the grand average legs for 29 April and 20 May in Fig. 8. However,  $\Delta LE / \Delta H > -1$  for the best-fit lines for both 29 April and 10 May, indicating that the smoothed aircraft fluxes do not match the surface-flux pattern. Also, considerable

TABLE 3. LSM standard deviations of  $H$ ,  $LE$ , and  $H + LE = R_n - G$  beneath the King Air flight tracks. LSM values smoothed to replicate aircraft four-point running averages.

Date	Model	$H$	$\sigma_H$	$LE$	$\sigma_{LE}$	$H + LE$	$\sigma_{H+LE}$
29 Apr	NCAR	235	28	89	31	324	14
	OSU	218	14	121	20	338	14
	SOLVEG	284	26	104	14	388	24
10 May	NCAR	176	40	270	59	446	22
	OSU	173	11	231	6	405	12
	SOLVEG	203	18	258	8	461	17
20 May	NCAR	165	6	308	6	472	7
	OSU	171	4	249	13	420	11
	SOLVEG	170	8	307	18	476	12

scatter remains. Both results suggest that more filtering is needed to remove the flux-concentrating effects of the atmosphere.

Figure 9 shows how the slope  $\Delta LE/\Delta H$  varies with the length of the running-mean filter. Most apparent in the figure is the already mentioned tendency for the slope to decrease with increased averaging. The correlation coefficient  $R$  between  $H$  and  $LE$  increases with the steepness of the slope and the reduction of scatter by smoothing. On 29 April and 10 May, the  $\Delta LE/\Delta H$  slope is more negative for the King Air data (track 1)

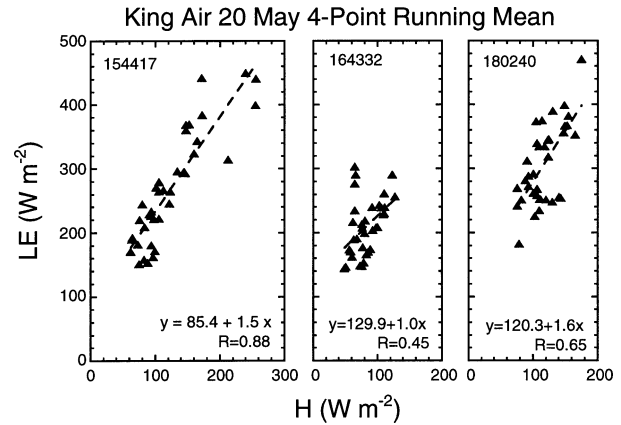


FIG. 7. As in Fig. 6 but for 20 May. Each point represents a four-point (3.6 km) running average.

than for the Twin Otter (track 3), though King Air  $R$  values exceed 0.5 on 29 April only. This is due both to the length of track 1, which allows more smoothing; and more along-track variability in  $H$  and  $LE$ , as suggested by Fig. 5. Only on 29 April do the aircraft data reproduce the slope  $\Delta LE/\Delta H \sim -1$ , for 12-point (11.6 km) running-mean averages. For this day, the negative

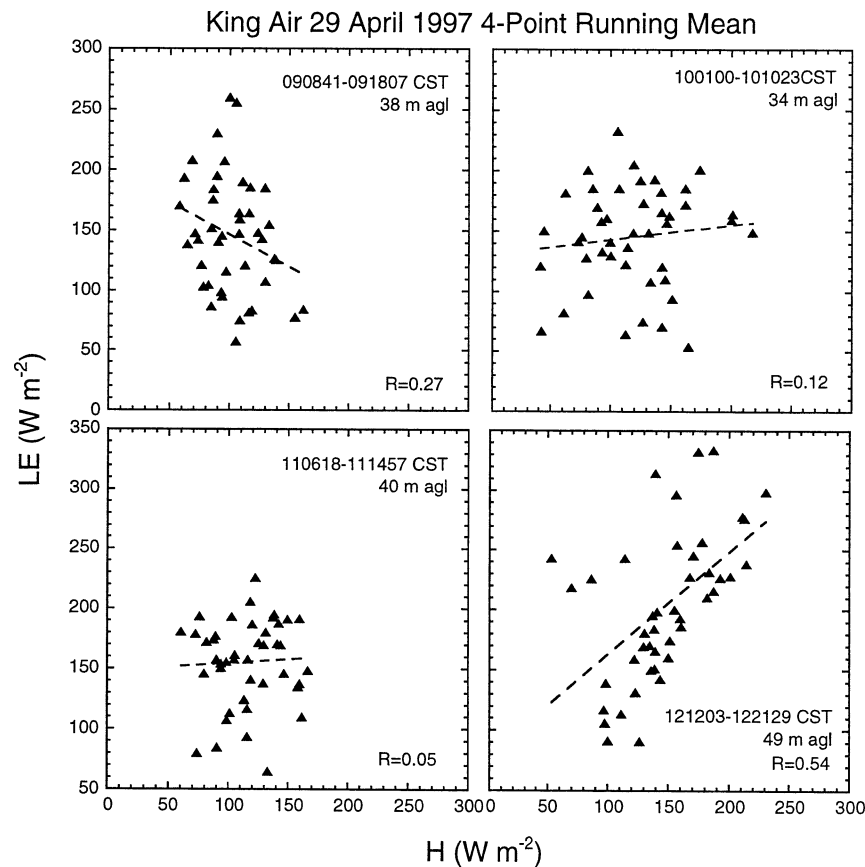


FIG. 6. For the morning of 29 Apr, the relationship of  $H$  and  $LE$  for the four King Air flight legs flown at 30 m. Each point represents a four-point (3.9 km) running average.



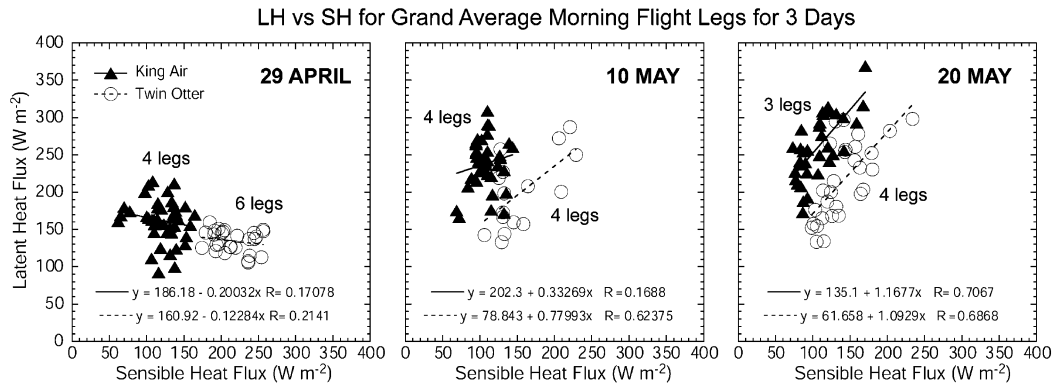


FIG. 8. The fluxes  $H$  vs  $LE$  for grand-average legs with four-point running averages for 29 Apr, 10 May, and 20 May. Each point represents  $0.01^\circ$  longitude along the leg; for length in km, see Fig. 2 and Table 2.

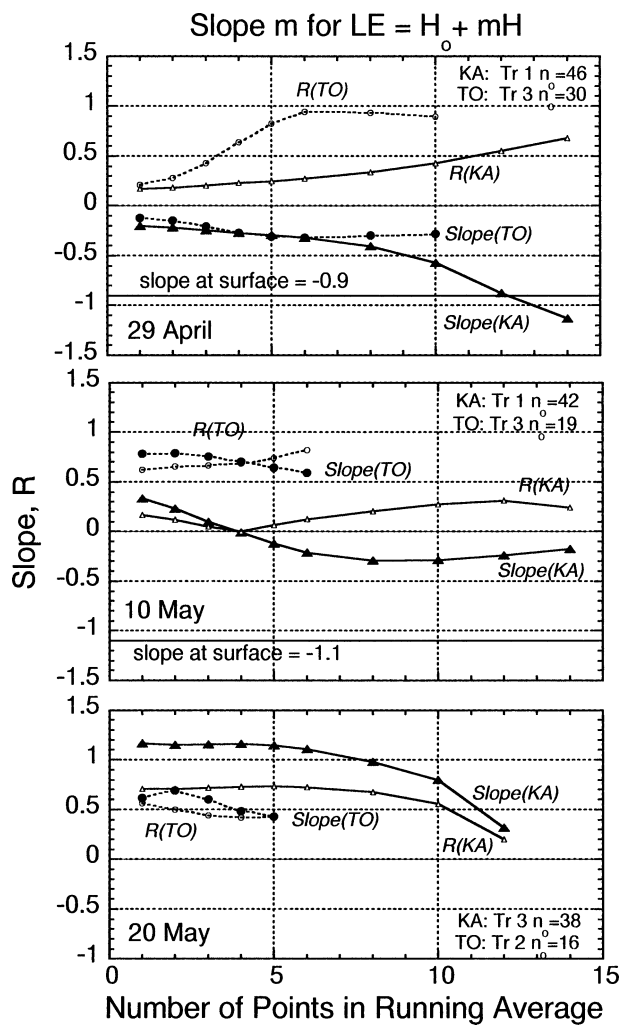


FIG. 9. Slope  $m$  of the least squares best-fit line  $LE = mH + H_o$  for 29 Apr, 10 May, and 20 May, for the King Air and Twin Otter, as a function of number of points in running mean. The total number of  $0.01^\circ$  segments in the grand average leg,  $n_o$ , is included. Results for filters longer than  $n_o/3$  points are not plotted since the equivalent number of block averages is less than three. For correspondence of number of points to distance, see Fig. 2.

slope for even two-point running averages indicates that some of the surface effect is evident. The slopes remain positive on 20 May for all averaging times.

*b. Statistical uncertainty from Mann and Lenschow (1994)*

We use (17) in Mann and Lenschow (1994) to estimate the fractional uncertainty for the average fluxes for the  $j$ th leg; that is,

$$\sigma_F^*(j) \equiv \frac{\sigma_F(j)}{F} = \left(\frac{2\lambda_f}{L_x}\right)^{1/2} \left(\frac{1 + r_{w's'}^2}{r_{w's'}^2}\right)^{1/2}, \quad (2)$$

where  $F$  is the flux averaged over the flight leg,  $L_x$  is the flight leg length,  $r_{w's'}$  is the correlation between  $w'$  and  $s'$ , and the integral scale  $\lambda_f$  is found from the frequency-weighted spectrum of the product  $w's'$  following Lenschow (1995) as described in LeMone et al. (2002). The resulting uncertainties for the individual low-level legs appear for the King Air in Table 4. In the table, the fractional uncertainty of the average flux for the grand average leg  $\sigma_F^*(ga)$  is approximated from the leg uncertainties  $\sigma_F^*(j)$  via

$$\sigma_F^*(ga) \approx \left[\sum_{j=1}^{nleg} \frac{1}{\sigma_F^{*2}(j)}\right]^{-1/2}, \quad (3)$$

where  $nleg$  is the number of legs. For the data here, (3) is nearly equivalent to taking the root-mean-square of  $\sigma_j^*$  and dividing by  $(nleg)^{1/2}$ .

From (2), the fractional uncertainty in fluxes  $\sigma_j^* \propto (1/L_x)^{1/2}$ , implying that the uncertainty of the grand average leg flux averaged spatially over  $n_{avg}$  points,  $\sigma_F^*(n_{avg})$ , is related to  $\sigma_F^*(ga)$  via

$$\sigma_F^*(n_{avg}) = \sigma_F^*(ga) \left(\frac{n_o}{n_{avg}}\right)^{1/2}, \quad (4)$$

where  $n_o$  is the number of points in the grand average leg before applying a running mean. The Twin Otter fractional uncertainties are calculated using the King Air

TABLE 4. Uncertainties for individual low-level legs [from (2)] and for grand average legs [from (3)]. Fractional uncertainties for grand average legs are used for Fig. 10.

Date	Time (CST)	$\overline{w'q'}$ (g kg <sup>-1</sup> m s <sup>-1</sup> )	$\sigma_{w'q'}^*$ (j)	$\sigma_{w'q'}^*$ (ga)	$\overline{w'\theta'}$ (K m s <sup>-1</sup> )	$\sigma_{w'\theta'}^*$ (j)	$\sigma_{w'\theta'}^*$ (ga)
29 Apr	0908:41–0918:07	0.049	0.15		0.087	15	
	1001:00–1010:23	0.055	0.19		0.096	15	
	1106:18–1114:57	0.056	0.15		0.103	10	
	1212:03–1221:29	0.070	0.21	0.09	0.132	11	0.06
10 May	0925:52–0935:12	0.077	0.15		0.067	20	
	1019:23–1029:34	0.064	0.19		0.086	22	
	1115:46–1125:02	0.078	0.16		0.099	24	
	1220:10–1230:07	0.101	0.15	0.08	0.100	24	0.11
20 May	0944:17–0952:25	0.090	0.15		0.103	10	
	1043:32–1051:06	0.074	0.16		0.072	12	
	1202:40–1210:44	0.104	0.17	0.09	0.100	12	0.07

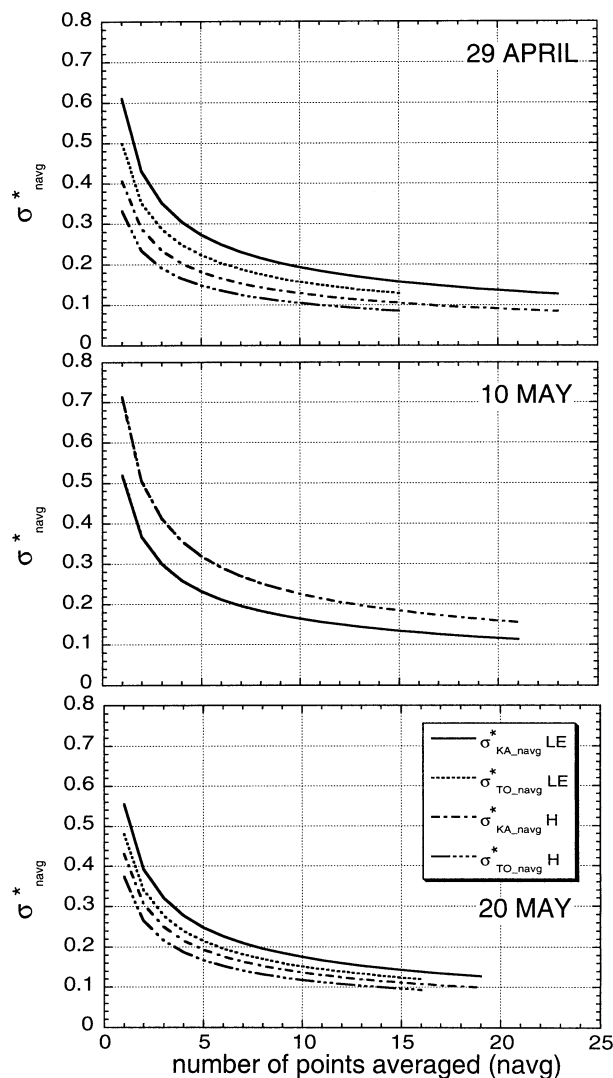


FIG. 10. Approximate uncertainty based on Table 4, Lenschow (1995), and ratio of averaging length to leg length (see text). For conversion of number of points to distance, see Table 2 and Fig. 2. For 10 May,  $\sigma_{\text{navg}}^*$  is the same for both aircraft. Values for Twin Otter, 20 May, were based on the four longer legs; so uncertainty estimates would be reduced  $\sim 10\%$  if all five legs were used.

$\sigma_F^*(\text{ga})$  values in Table 3 and adjusting for the differences in total record length for the two aircraft; that is,

$$\sigma_F^*(\text{navg}_{\text{TO}}) = \sigma_F^*(\text{ga}) \left( \frac{n_{o,\text{TO}}}{\text{navg}_{\text{TO}}} \right)^{1/2} \left( \frac{n_{o,\text{KA}} \times \text{nleg}_{\text{KA}}}{n_{o,\text{TO}} \times \text{nleg}_{\text{TO}}} \right)^{1/2}, \quad (5)$$

where  $(n_{o,\text{KA}} \times \text{nleg}_{\text{KA}})/(n_{o,\text{TO}} \times \text{nleg}_{\text{TO}}) \approx L_x(\text{total, King Air})/L_x(\text{total, Twin Otter})^3$  and subscripts KA and TO refer to King Air and Twin Otter respectively.

Figure 10 shows the fractional uncertainties from Table 3, (4), and (5) as a function of running-mean filter length, for the two aircraft. For the King Air, an  $H$  uncertainty of 20% can be achieved for four-point averages on 29 April and five-point averages on 20 May, but a large flux integral scale for  $w'\theta'$  on 10 May leads to a requirement of 13-point averages for similar accuracy. A 20% accuracy for LE requires 10-point averaging for 29 April, 7-point averaging for 10 May, and 8-point averaging for 20 May.

### c. Evaluating spatial and temporal standard deviations

Low-pass filtering at longer wavelengths increases statistical confidence and improves the adherence of the aircraft  $H$  and LE fluxes to the surface-energy budget constraint, but it also reduces the horizontal variability we need to compare with LSMs. In this section, we examine the effect of low-pass filtering on the spatial variability of the grand average leg. We will also compute temporal variability centered about each point as a function of filter length and compare this to the corresponding spatial variability, to see if there are filter lengths for which spatial variability approaches or exceeds temporal variability. Maximizing spatial variability compared to variability at a point (temporal variability) increases the chance of statistical confidence in tests of the robustness of extrema.

<sup>3</sup> Since the basic segment length is  $0.01^\circ$  longitude, the relationship is exact only if the two aircraft tracks are exactly parallel.

1) SPATIAL STANDARD DEVIATION,  $\sigma_{\text{spat}}(\text{ga})$

After low-pass filtering with the running mean filter, we obtain the spatial standard deviation for the grand average leg  $\sigma_{\text{spat}}(\text{ga})$  from

$$\sigma_{\text{spat}}^2(\text{ga}) = \frac{1}{n} \sum_{i=1}^n [F_{\text{ga}}(i) - F_{\text{ga}}]^2, \quad (6)$$

where  $n$  is the number of points in the leg;  $F_{\text{ga}}(i)$  is the grand average leg flux for the line segment centered at point  $i$  and consisting of  $\text{navg}$  points, given by

$$F_{\text{ga}}(i) = \frac{1}{\text{nleg}} \sum_{j=1}^{\text{nleg}} F(i, j),$$

where  $F(i, j)$  is the flux for line segment  $i$  on the  $j$ th leg; and  $F_{\text{ga}}$  is the grand average leg average flux,

$$F_{\text{ga}} = \frac{1}{n} \sum_{i=1}^n F_{\text{ga}}(i).$$

2) TEMPORAL STANDARD DEVIATION,  $\sigma_t(i)$

(a) Method 1

Calculate standard deviations relative to time average at each point  $i$ . That is,

$$\sigma_{t1}^2(i) = \frac{1}{\text{nleg}} \sum_{j=1}^{\text{nleg}} [F(i, j) - F_{\text{ga}}(i)]^2. \quad (7)$$

Because there is a time trend (Table 4),  $\sigma_{t1}(i)$  is nonzero even for no random temporal variability. For example, if the flux for the  $i$ th segment is  $F = a + bt$ , where  $t = \text{time}$ ,  $\sigma_{t1}(i) = bt/(2\sqrt{3})$ .

(b) Method 2

This method is designed to remove the effect of the time trend before calculating the temporal standard deviation, to approximate the random temporal standard deviation. First, we find the time trend  $\text{Tr}$  using a least squares linear fit to the mean leg fluxes  $F(j)$  plotted as a function of time. Second,  $F'(i, j)$  is calculated for the  $i$ th segment of leg  $j$  by subtracting the flux predicted for that segment from  $\text{Tr}$ :

$$F'(i, j) = F(i, j) - \text{Tr}(i, j), \quad (8a)$$

where  $\text{Tr}(i, j)$  is evaluated at the center time of the segment. Before calculating the temporal standard deviation, we need to calculate and then remove the mean departure from the linear trend for the  $i$ th segment,  $F'_{\text{ga}}(i)$ , since this contains the ‘‘signal’’ we are looking for:

$$F'_{\text{ga}}(i) = \frac{1}{\text{nleg}} \sum_{j=1}^{\text{nleg}} F'(i, j) \quad (8b)$$

$$\sigma_{t2}^2(i) = \frac{1}{\text{nleg}} \sum_{j=1}^{\text{nleg}} [F'(i, j) - F'_{\text{ga}}(i)]^2. \quad (8c)$$

Note that without a linear trend,  $F'(i, j) \equiv F(i, j)$ ,  $F'_{\text{ga}}(i) \equiv F_{\text{ga}}(i)$ , and (7) and (8c) are equivalent.<sup>4</sup> We also applied method 2 using the leg-center-time value of  $\text{Tr}$  to all segments in each leg, but the difference in the results was small, since the leg duration (maximum 9 min; Table 4) is short compared to the time over which measurements were taken (4 h).

Figures 11 and 12 compare the two measures of temporal variability  $\sigma_{t1}(i)$  (error bars) and  $\sigma_{t2}(i)$  (dashed line) to the spatial variability  $\sigma_{\text{spat}}$  for a four-point running average. Removal of the time trend before calculating the standard deviation reduces the temporal variability slightly for this dataset. The temporal variability exceeds the spatial variability at most of the points. The spatially averaged temporal variability for the grand average leg, is given by

$$\sigma_{tk} = \frac{1}{n} \sum_{i=1}^n \sigma_{ik}(i), \quad (9)$$

where  $k = 1$  or  $2$ . We will use  $\sigma_{ik}$  when comparing spatial to temporal variability for different filtering lengths.

Figures 13 and 14 compare the spatial standard deviation for the grand average leg  $\sigma_{\text{spat}}(\text{ga})$  (henceforth called  $\sigma_{\text{spat}}$  for brevity) to the spatially averaged temporal standard deviations  $\sigma_{t1}$  and  $\sigma_{t2}$ . The figures show the effect of horizontal filtering on both spatial and temporal variability. Ideally, we want to minimize the ‘‘random’’ variability  $\sigma_{t2}$  relative to  $\sigma_{\text{spat}}$  while accounting for statistical uncertainty (Fig. 10) and reproducing the  $\Delta\text{LH}/\Delta H$  slope defined by the surface observations.

Both spatial and temporal standard deviations decrease sharply as the running-mean filter length changes from 2 to 10 points. For shorter-length running-mean filters, the fluxes have larger statistical uncertainty (see Fig. 10), and both temporal and spatial variability probably reflect individual flux-transport events in the atmosphere in addition to the sought-after effect of surface fluxes as noted previously. As the filter wavelength increases, the spatial variability begins to be filtered out, and the temporal standard deviations reflect the scatter of the average leg fluxes about the mean or linear trend. The average temporal standard deviation for the grand average leg almost always exceeds the spatial standard deviation. On 29 April for the *King Air*,  $\sigma_{\text{spat}}$  approaches  $\sigma_{t2}$  for LE and actually exceeds  $\sigma_{t2}$  for *H* at filter wavelengths between four and nine points. This is also a reasonable interval for obtaining reasonable statistics (Fig. 10) and adherence to the surface energy budget constraint (Fig. 9). For the *Twin Otter*, the spatial variability exceeds  $\sigma_{t2}$  for *H* at nearly all filter wavelengths on 10 May.

<sup>4</sup> An alternative way to remove the time trend from the temporal standard deviation would be to subtract out the leg means from  $F(i, j)$  and then calculate the temporal standard deviation of the resulting fluctuations. This is equivalent to method 2 if the leg averages lie on a straight line.

4-point Running Average Fluxes with  $\sigma_{temporal}$

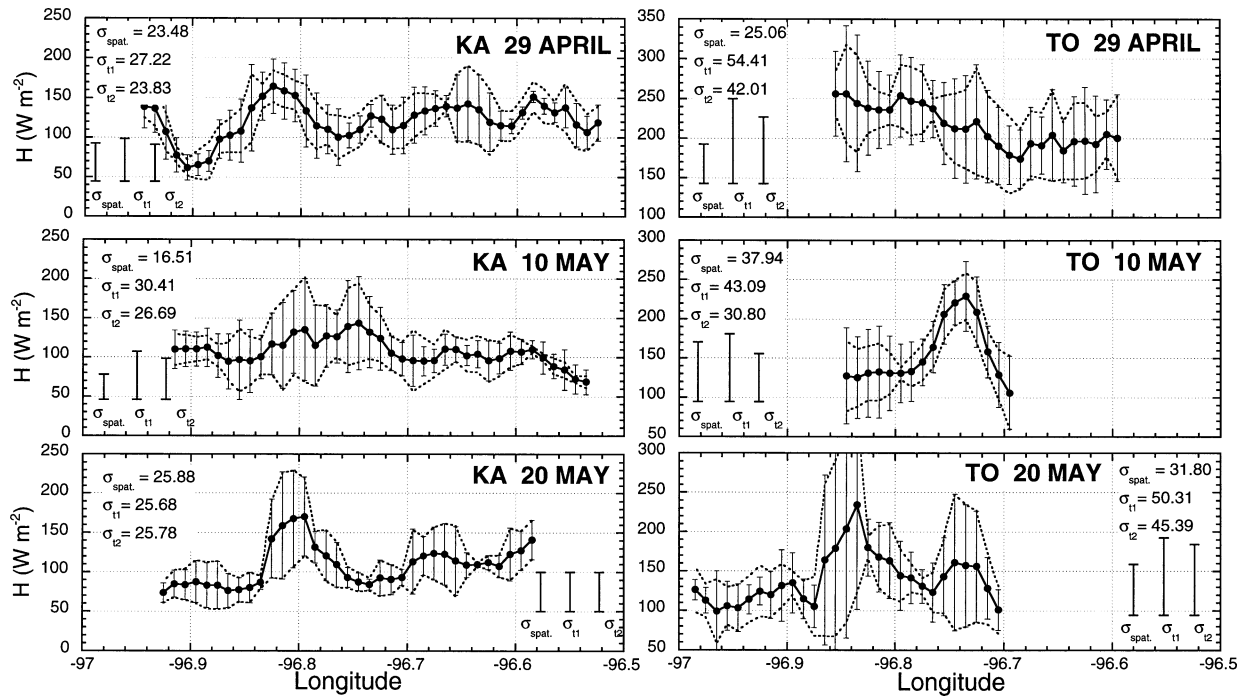


FIG. 11. For four-point running averages,  $H$  as a function of longitude. Standard deviations are based on method 1 ( $\sigma_{t1}$ , error bars) and method 2 ( $\sigma_{t2}$ , dashed lines). Spatial standard deviation  $\sigma_{spat}$  is included for comparison. For conversion of longitude and averages to km, see Table 2 and Fig 2.

4-point Running Average Fluxes with  $\sigma_{temporal}$

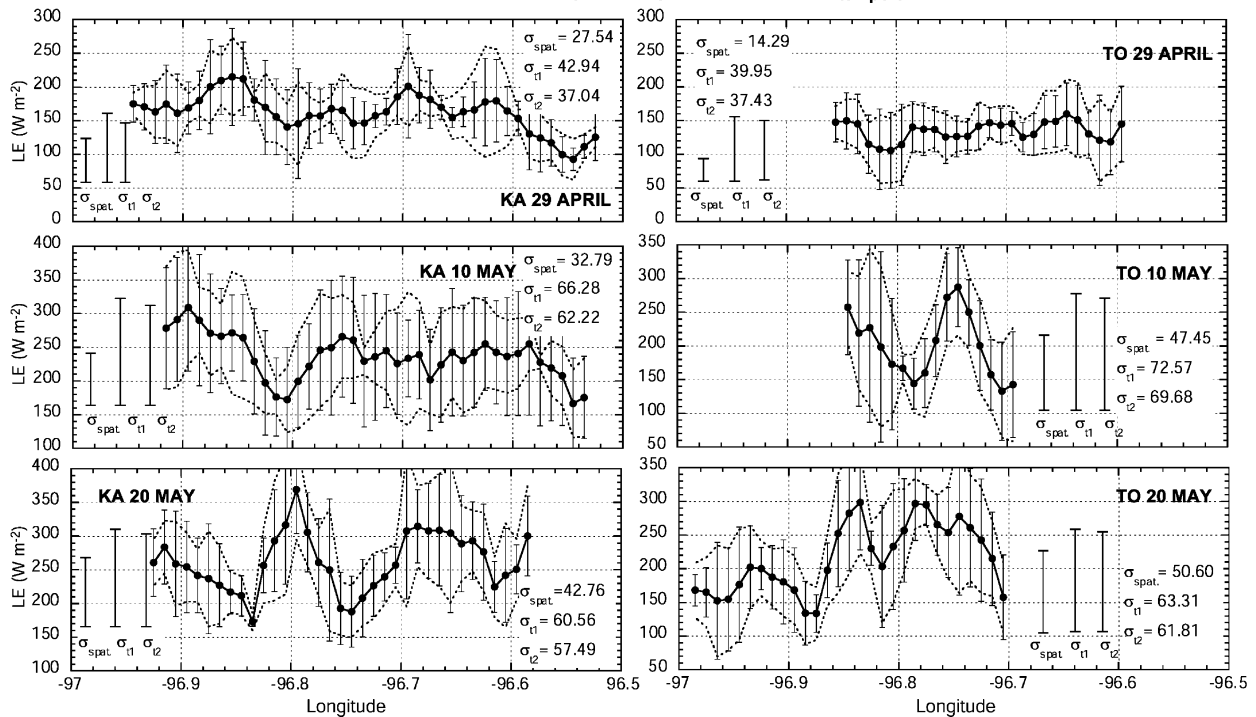


FIG. 12. As in Fig. 11, but for LE.

Wyoming King Air - Effect of Running Average

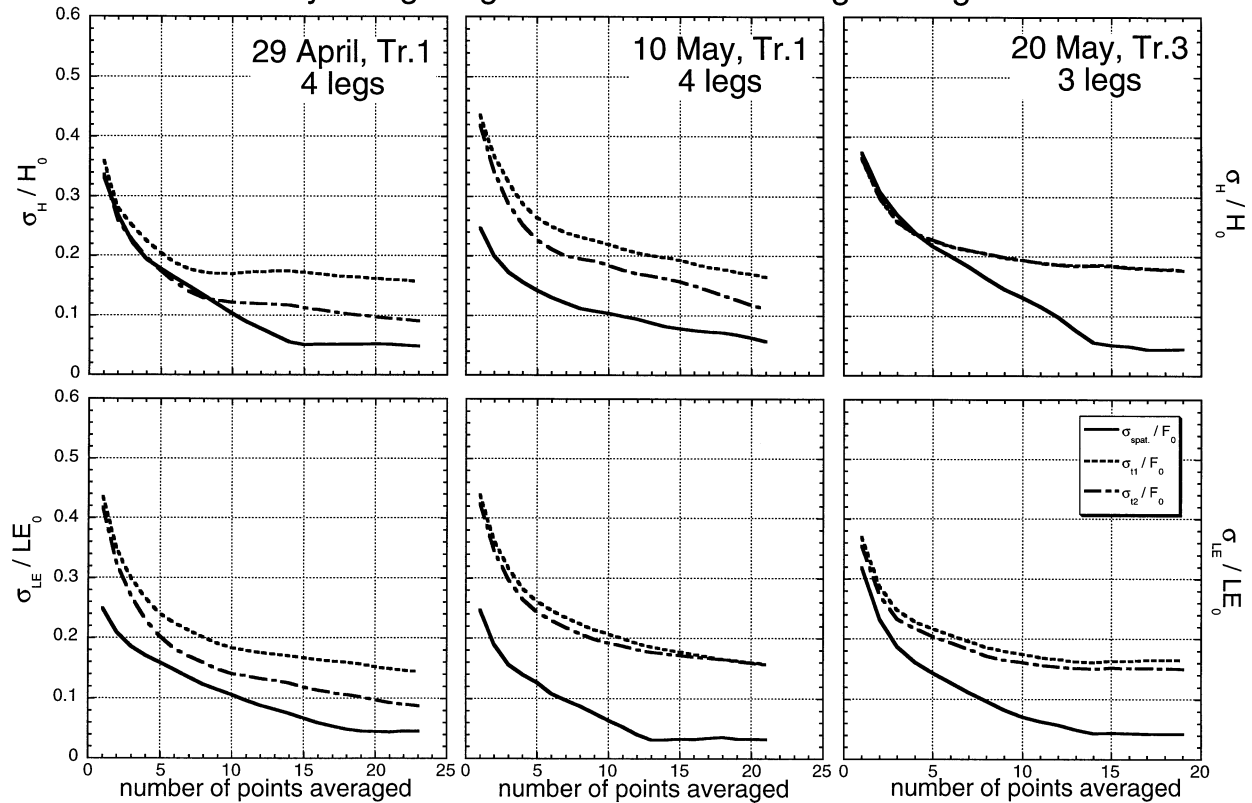


FIG. 13. Comparison of grand average leg spatial standard deviation (solid line) to the average temporal standard deviation for the King Air, as a function the number of points of the running average filter, based on method 1 (dash line) and method 2 (long-short dash line). Standard deviations are normalized by the grand average leg average flux before smoothing. For conversion of points averaged to km, see Fig. 2.

4. Synthesis

To illustrate the application of the three steps taken in evaluating the CASES-97 flux data, we apply them to the three days separately.

a. 29 April

This day has the largest horizontal variability in fluxes of the three days (Figs. 1, 4), which is sufficient to produce a negative  $H$ - $LE$  slope for one flight leg in Fig. 6 (though not statistically significant:  $R$  only 0.27); and the slope for the grand average legs is the least positive for the three days (Fig. 8). However, the slope is near zero and ill defined, indicating a need for more horizontal averaging. Filtering is sufficient to make the aircraft  $H$ - $LE$  slope match that at the surface on 29 April (the one day of the three) for King Air 12-point (11.6 km) averages (Fig. 9). For the King Air, average random temporal variability ( $\sigma_2$ , trend removed) and spatial variability for  $H$  are comparable for short spatial averages (<10 points; Fig. 13), but there is less horizontal variability for the Twin Otter (Fig. 14). Four-point running means are sufficient for 20% uncertainty in  $H$ , but 10 points are required for 20% uncertainty in  $LE$  for the

King Air; with fewer points (3 for  $H$  and 7 for  $LE$ ) required for the Twin Otter because it flew more legs (Table 2).

In Fig. 15,  $H$  and  $LE$  are shown for the grand average legs for 4-, 8-, and 12-point running mean filters, to bracket the averaging intervals that appear most desirable. From the figure, the 12-point filter is excessive. Most of the horizontal variability is removed by 12-point averaging, as expected from Fig. 13. Two  $H$  features that look robust with the lighter filtering—the minimum near the west end of the track with temporal standard deviations smaller than the spatial standard deviation for both the 4-point and 8-point filters, and a minimum near the center of the track that looks robust for 8-point filtering—are less distinct for the 12-point filtering, again as expected from Fig. 13. Further, with  $R$  so small (Fig. 9), there is uncertainty in the aircraft  $\Delta LE/\Delta H$  slopes. Thus it seems reasonable to choose a running-mean filter between four and eight points, and accept that some of the spatial variability might not be replicated by the models. This is the choice made by Chen et al. (2003).

If the experiment could be repeated, we would want to fly more flight tracks instead of filtering over more

NOAA Twin Otter - Effect of Running Average

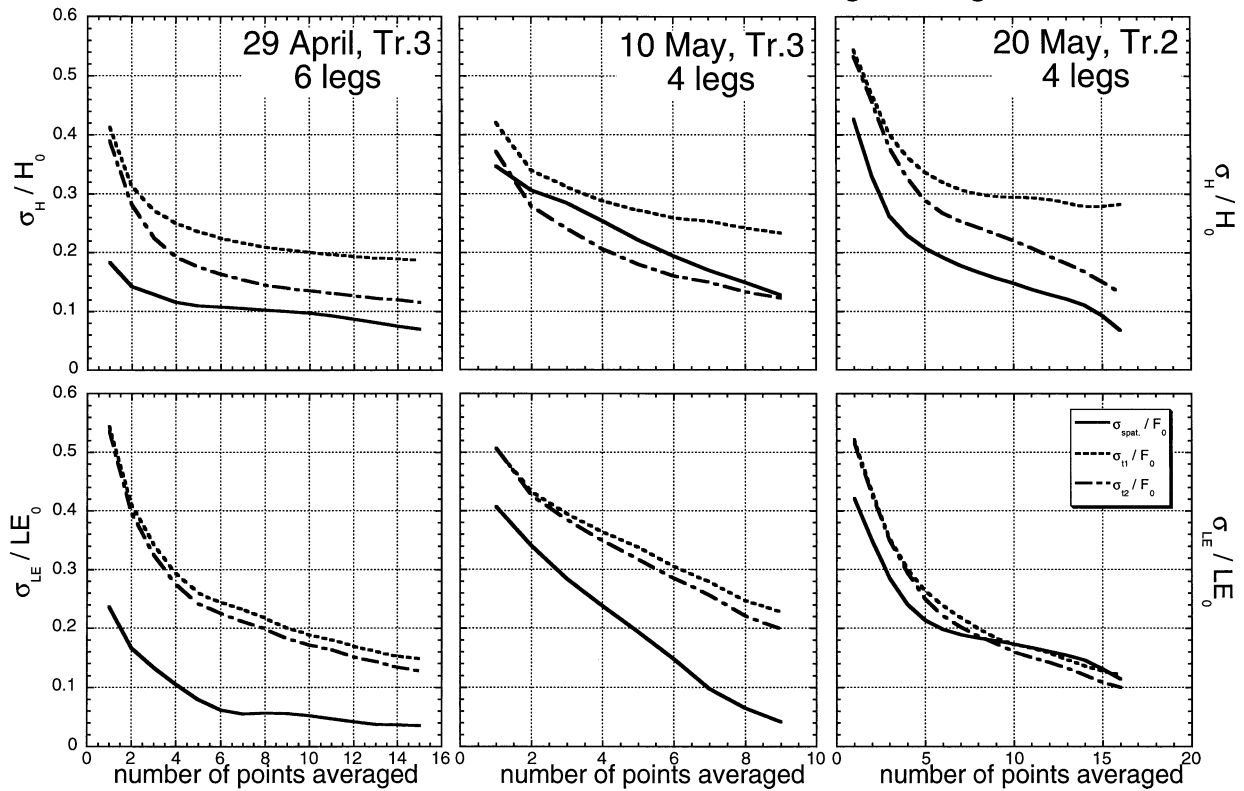


FIG. 14. As in Fig. 13, but for the Twin Otter.

Wyoming King Air 29 APRIL

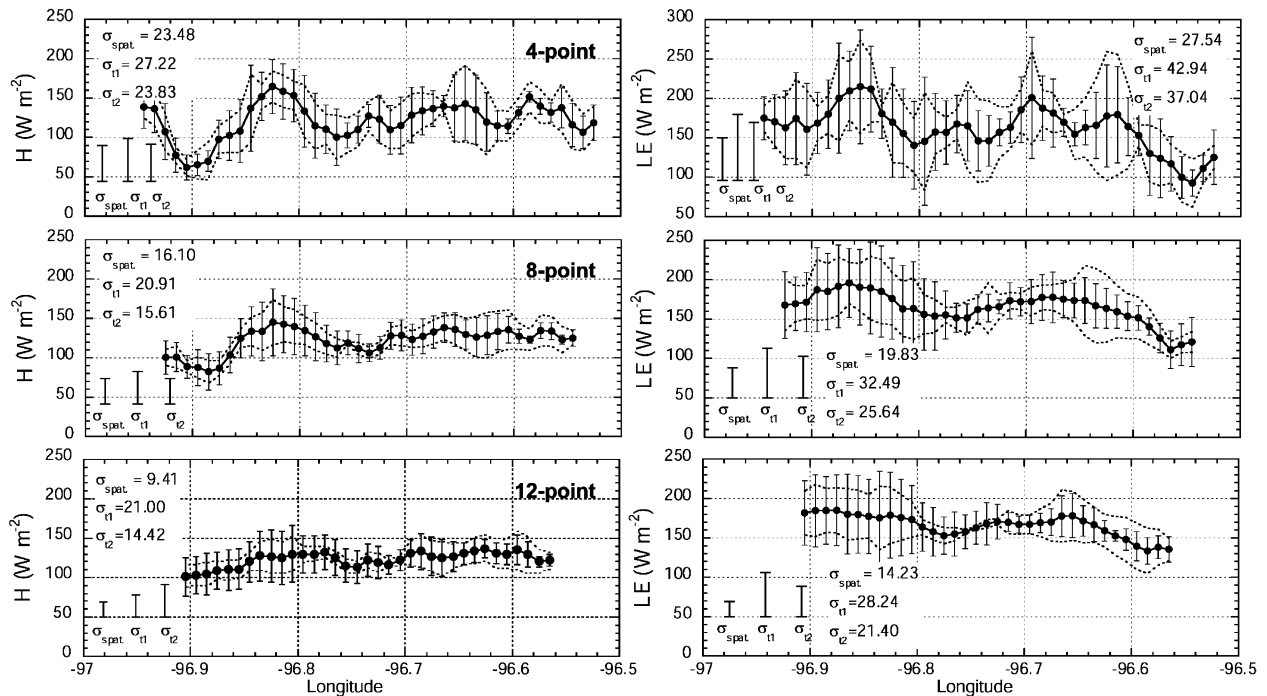


FIG. 15. For 29 Apr grand average leg,  $H$  and  $LE$  for 4-point (3.9 km), 8-point (7.8 km), and 12-point (11.6 km) filters. Error bars as in Fig. 11.

points. Assuming that increasing the sample to the desired size by changing the number of flight legs (and keeping the filter length constant) is equivalent to changing the sample by increasing the filter length (and keeping the number of legs constant), we would need to fly 12 low-level legs to reproduce the  $-0.9$  slope with a four-point filter for the King Air.

#### b. 10 May

Both the observed (Fig. 4) and modeled (Fig. 5) horizontal variability are intermediate between those of 29 April and 20 May, because of a smaller range in sensible heat flux. Statistical uncertainty in  $H$  is larger than that for the other two days (Fig. 10), due to the large integral scale for  $w'\theta'$  (LeMone et al. 2002). Consistent with the smaller spread for  $H$  in the surface data (Fig. 4), the temporal variability for the King Air exceeds the spatial variability significantly for nearly all averaging times (Fig. 13). However, the Twin Otter data show the opposite result (Fig. 14) because of significant spatial variability in the aircraft fluxes along track 3 (Fig. 11; this is not supported by the LSMs in Fig. 5, which show less horizontal variability for the the Twin Otter track compared to the King Air track). Uncertainty in LE is comparable to that for 29 April (Fig. 10). Perhaps related to the larger  $H$  uncertainty compared to 29 April, the aircraft-derived  $H$ -LE slope becomes only slightly negative, with the most negative value of  $\Delta LE/\Delta H$  for an 8–10-point filter on King Air track 1 (Fig. 9). Chen et al. (2003) show better agreement between predicted and observed variability for the King Air than for the Twin Otter, as would be expected from the slope behavior. However, model-aircraft agreement on 10 May is no worse than that for 29 April, contrary to expectations based on slope behavior and statistical uncertainty.

#### c. 20 May

From a purely statistical standpoint, Fig. 10 shows that the data on 20 May has the least overall uncertainty. However, the surface conditions are so homogeneous that even the surface data do not show a negative correlation between  $H$  and LE (Fig. 4), even though the land surface models do (Fig. 5). Despite the surface uniformity, the horizontal variability in the aircraft fluxes is comparable to that on the more heterogeneous days; but this reflects concentration of fluxes by atmospheric processes: the aircraft data have a positive  $H$ -LE slope for all filter lengths (Fig. 9). Without a surface pattern to follow, it is difficult to define a “correct” slope, so we have to rely on the other two factors to judge the quality of the data. However, the leg-averaged fluxes would be excellent for comparison to models. While most comparisons in Chen et al. (2003) for this day are extremely poor, the comparison of LE variability to the NCAR LSM and the atmosphere-soil-vegetation model SOLVEG are marginally better than for the other days.

## 5. Discussion and conclusions

We have described three factors to consider in choosing the best horizontal-averaging length for using aircraft data to evaluate the success of LSMs in replicating horizontal variability of sensible and latent heat fluxes. Using fluxes averaged with running-mean filters of varying lengths, we have 1) examined the similarity of the relationship between  $H$  and LE at the surface and flight level (the “surface energy budget constraint”), 2) estimated the percentage uncertainty from Mann and Lenschow (1994), and 3) compared spatial and temporal variability.

First, we examine adherence of aircraft-derived fluxes to the surface energy budget constraint, that the  $\Delta LE/\Delta H$  slope is negative for the best-fit line on a plot of  $H$  versus LE for a given time if LE or  $H$  vary more than  $R_n - G$  (slope =  $-1$  if  $R_n - G \approx \text{constant}$ ). If the observed surface fluxes meet this constraint, the aircraft data also must meet this constraint for a meaningful comparison to LSMs. The CASES-97 surface data meet the constraint for days with large horizontal variability, but the negative-slope criterion was not useful for 20 May, because there was not enough horizontal variability to determine a slope. Applying a longer-wavelength filter to the aircraft data improves the chances of the right slope, but at the possible expense of losing horizontal variability.

In practice, significant horizontal variability in the surface values of  $H$  and LE is required for this constraint to be met, and the aircraft should be flying at a height low enough that systematic height changes in  $H$  and LE do not significantly affect their relative magnitudes. Furthermore, since the filtering is designed to remove the effects of flux concentration by large eddies, this technique will not work if the large eddies are stationary. Large eddies in the convective boundary layer tend to move with the mean boundary layer wind. Thus long-lived large eddies in calm winds could produce average flux extremes not associated with surface heterogeneity. Roll vortices could cause a similar problem in stronger winds, since rolls are quasi-two-dimensional and are nearly parallel to the wind. However, all of the rolls documented by LeMone (1973) traveled across the surface at a rate of about  $1 \text{ m s}^{-1}$ , giving typical advection times of 0.5–1 h. Finally, terrain could lead to preferred large-eddy locations or concentrate fluxes in other ways.<sup>5</sup>

Second, we estimate the statistical uncertainty for the running average flux on the grand average leg following Mann and Lenschow (1994), by multiplying the uncertainty for the grand average leg flux by the square root of the ratio of the grand average leg length to the filter length. Since uncertainty and horizontal variability in

<sup>5</sup> One of the authors (R. Grossman) reported persistent fluxes at some locations in CASES-97 in the December 2000 American Geophysical Union meeting; this is the subject of a future paper.

$H$  and LE are both functions of filter length, acceptable flux accuracy has to be balanced against variability.

Finally, we compare temporal and spatial variability as a function of filter length. Other things being equal, the best filter length range is that for which the spatial variability in the grand average leg is largest compared to random temporal variability but still admits enough horizontal variability for meaningful comparisons to LSMs. Too short a filter length leaves in variability related to atmospheric events (which leads to positive correlation between  $H$  and LE); too long a filter length and the horizontal variability is filtered out.

Thus the ideal dataset for comparison of observed to modeled variations in  $H$  and LE along a flight track would have the following attributes. At the surface, there would be a range of sensible and latent heat fluxes, varying on scales large enough to be resolved after spatial filtering, and a well-defined negative slope on a plot of  $H$  versus LE for a given time, giving us a tool for knowing when the atmospheric flux concentration effect is removed via filtering from the aircraft fluxes. For the aircraft data, we look for adherence to the surface  $H$ –LE relationship, small statistical uncertainty, and significant spatial variability that is either close to or greater than random temporal variability. We have shown that the aircraft data can be made to meet these criteria separately by filtering. However, filtering sufficient for a tolerable statistical uncertainty or to obtain the right  $H$ –LE slopes can also reduce the horizontal variability to unacceptable levels.

Applying these steps to the Chen et al. (2003) analysis, the data collected on 29 April was the most suitable for LSM evaluation. Even for this day, agreement was only fair. In future comparisons, it is important to fly as many legs as possible over a track with as much heterogeneity as possible, on scales resolvable after whatever filtering necessary to simultaneously replicate an acceptable statistical uncertainty and  $H$ –LE slope. Based on the analysis here, 12 flight legs over track 1 would have provided an optimum comparison. In addition, recent work by J. Song and M. Wesely (2002, personal communication) suggest that comparing the aircraft data to the fetch upstream following the approach of Schuepp et al. (1992) will also improve the consistency of aircraft and surface fluxes. Finally, it would be interesting to see whether high-pass filtering the data before estimating fluxes would eliminate some of the large-eddy effects and improve correlations with model results.

*Acknowledgments.* The authors wish to acknowledge the input from Joe Alfieri and three anonymous reviewers, who improved the manuscript and made it more palatable for the surface–atmosphere interaction community. This work would not have been possible without the skill and dedication of the scientific and operations crews for the University of Wyoming *King Air* and the NOAA *Twin Otter*, and the scientific and technical crew

at NCAR, who maintained the surface-flux stations. RLG's work was supported by NSF Grant ATM-9615583; DY and FC's work was supported by the NASA Land-Surface Hydrology Program under Award NAG5-7593.

#### REFERENCES

- Bonan, G. B., 1996: A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and users guide. NCAR Tech. Note NCAR/TN-417+STR, 150 pp. [Available from NCAR Library, P.O. Box 3000, Boulder, CO 80307.]
- Chen, F., and Coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, **101**, 7251–7268.
- , D. N. Yates, H. Nagai, M. A. LeMone, K. Ikeda, and R. L. Grossman, 2003: Land surface heterogeneity in the Cooperative Atmosphere Surface Exchange Study (CASES-97). Part I: Comparison of modeled surface flux maps with surface-flux tower and aircraft measurements. *J. Hydrometeorol.*, **4**, 196–218.
- Deardorff, J., 1974: Three-dimensional numerical study of turbulence in an entraining mixed layer. *Bound.-Layer Meteorol.*, **32**, 205–236.
- Desjardins, R. L., P. H. Schuepp, J. I. MacPherson, and D. J. Buckley, 1992: Spatial and temporal variations of the fluxes of carbon dioxide and sensible and latent heat over the FIFE site. *J. Geophys. Res.*, **97** (D17), 18 467–18 475.
- Fritschen, L., and P. Qian, 1992: Variation in energy balance components from six sites in the native prairie for three years. *J. Geophys. Res.*, **97** (D17), 18 651–18 661.
- Grossman, R. L., D. N. Yates, M. A. LeMone, and K. Ikeda, 2000: Land-surface features associated with aircraft observations of variations in sensible heat flux: Further evidence of the effects of large eddies on surface fluxes. *Eos, Trans. Amer. Geophys. Union*, **81** (48), F148.
- Horst, T. W., and J. C. Weil, 1992: Footprint estimation for scalar flux measurements in the atmospheric surface layer. *Bound.-Layer Meteorol.*, **59**, 279–296.
- , and S. P. Oncley, 1995: Flux-PAM measurement of scalar fluxes using cospectral similarity. Preprints, *Ninth Symp. on Meteorological Observations and Instrumentation*, Charlotte, NC, Amer. Meteor. Soc., 495–500.
- Hunt, J. C. R., J. C. Kaimal, and J. E. Gaynor, 1988: Eddy structure in the convective boundary layer—New measurements and new concepts. *Quart. J. Roy. Meteor. Soc.*, **114**, 827–858.
- Kaimal, J. C., J. C. Wyngaard, D. A. Haugen, O. R. Cote, Y. Izumi, S. J. Caughey, and C. J. Readings, 1976: Turbulence structure in the convective boundary layer. *J. Atmos. Sci.*, **33**, 2152–2169.
- Kelly, R. D., E. A. Smith, and J. I. MacPherson, 1992: A comparison of surface sensible and latent heat flux from aircraft and surface measurements in FIFE 1987. *J. Geophys. Res.*, **97** (D17), 18 445–18 453.
- Khalsa, S. J. S., and J. A. Businger, 1977: The drag coefficient as determined by the dissipation method and its relation to intermittent convection in the surface layer. *Bound.-Layer Meteorol.*, **12**, 273–297.
- Lambert, D., P. Durand, F. Thoumieux, B. Benech, and A. Druilhet, 1999: The marine atmospheric boundary layer during SEMAPHORE II: Turbulence profiles in the mixed layer. *Quart. J. Roy. Meteor. Soc.*, **125**, 513–528.
- LeMone, M. A., 1973: The structure and dynamics of horizontal roll vortices in the planetary boundary layer. *J. Atmos. Sci.*, **30**, 1077–1091.
- , 1976: Modulation of turbulence energy by longitudinal rolls in an unstable planetary boundary layer. *J. Atmos. Sci.*, **33**, 1308–1320.
- , 1990: Some observations of vertical velocity skewness in the



- convective planetary boundary layer. *J. Atmos. Sci.*, **47**, 1163–1169.
- , and W. T. Pennell, 1976: The relationship of trade-wind cumulus distribution to subcloud-layer fluxes and structure. *Mon. Wea. Rev.*, **104**, 524–539.
- , and Coauthors, 2000: Land–atmosphere interaction research and opportunities in the Walnut River watershed in southeast Kansas: CASES and ABLE. *Bull. Amer. Meteor. Soc.*, **81**, 757–780.
- , and Coauthors, 2002: CASES-97: Late-morning warming and moistening of the convective boundary layer over the Walnut River watershed. *Bound.-Layer Meteor.*, **104**, 1–52.
- Lenschow, D. H., 1995: Micrometeorological techniques for measuring biosphere–atmosphere trace gas exchange. *Biogenic Trace Gases: Measuring Emissions from Soil and Water*, P. Matson and R. Harriss, Eds., Blackwell Science, 126–163.
- , J. C. Wyngaard, and W. T. Pennell, 1980: Mean-field and second-moment budgets in a baroclinic, convective boundary layer. *J. Atmos. Sci.*, **37**, 1313–1326.
- Mahrt, L., 1991: Boundary-layer moisture regimes. *Quart. J. Roy. Meteor. Soc.*, **117**, 151–176.
- , 1998: Flux sampling errors for aircraft and towers. *J. Atmos. Oceanic Technol.*, **15**, 416–429.
- Mann, J., and D. H. Lenschow, 1994: Errors in airborne flux measurements. *J. Geophys. Res.*, **99**, 14 519–14 526.
- Millitzer, J. M., M. C. Mihaelis, S. R. Semmer, K. S. Norris, T. W. Horst, S. P. Oncley, A. C. Delany, and F. C. Brock, 1995: Development of the prototype PAM III/Flux-PAM surface meteorological station. Preprints, *Ninth Symp. on Meteorological Observations and Instrumentation*, Charlotte, NC, Amer. Meteor. Soc., 490–494.
- Moeng, C.-H., and P. P. Sullivan, 1994: A comparison of shear- and buoyancy-driven planetary boundary layer flows. *J. Atmos. Sci.*, **51**, 999–1022.
- Nagai, H., 2002: Validation and sensitivity analysis of a new atmosphere–soil–vegetation model. *J. Appl. Meteor.*, **41**, 160–176.
- Pan, H.-L., and L. Mahrt, 1987: Interaction between soil hydrology and boundary-layer development. *Bound.-Layer Meteor.*, **38**, 185–202.
- Schuepp, P. H., J. I. MacPherson, and R. L. Desjardins, 1992: Adjustment of footprint correction for airborne flux mapping over the FIFE site. *J. Geophys. Res.*, **97** (D17), 18 455–18 466.
- Sellers, P. J., F. G. Hall, G. Asrar, D. E. Strebel, and R. E. Murphy, 1992: An overview of the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE). *J. Geophys. Res.*, **97**, 18 345–18 373.
- Yates, D. N., F. Chen, M. A. LeMone, R. Qualls, S. P. Oncley, R. L. Grossman, and E. A. Brandes, 2001: A Cooperative Atmosphere–Surface Exchange Study (CASES) dataset for analyzing and parameterizing the effects of land surface heterogeneity on area-averaged surface heat fluxes. *J. Appl. Meteor.*, **40**, 921–937.
- Young, G. S., 1988: Turbulence structure of the convective boundary layer. Part I: Variability of normalized turbulence statistics. *J. Atmos. Sci.*, **45**, 719–726.