Timescales and structures in vertical transport in the atmospheric boundary layer

Jerome Schalkwijk, Fred Bosveld and Pier Siebesma

Authors: Schalkwijk, J.
            Bosveld, F.C.
            Siebesma, A.P.

● ● ● ●

Timescales and Structures in Vertical Transport in the Atmospheric Boundary Layer

Jerome Schalkwijk   Fred Bosveld   Pier Siebesma

March 30, 2010

Abstract

A combination of high frequency measurements from the Cabauw measurement site and LES results is used to evaluate timescales and structures of turbulent transport of heat in the atmospheric boundary layer.

On the basis of the cospectrum of vertical velocity and temperature, a dominant timescale of transport is introduced. An emperical equation is proposed to easily determine this timescale as a function of wind velocity, height and Obukhov length. Using this timescale, loss in Eddy Covariance measurements due to finite averaging time can be predicted on the basis of a model for the cospectrum shape. Eddy Covariance measurements from the Cabauw site show behaviour confirming this predicted loss behaviour. No evidence is found for the possible presence of unpredicted systematic flux on very large timescales.

Attention is also given to the presence of a certain gap scale seperating turbulent and mesoscale motions, but it is found that these motions overlap in unstable conditions, making seperation impossible.

As several authors have pointed at the phenomena of Turbulent Organized Structures as large contributors to large timescale flux, an attempt is made to identify these structures in measurement series. This is done with the help of two algorithms, based on criteria and wavelet analysis, demanding vertical coherence and rising, warm air. Detected structures within measured and LES simulated time series show great similarity, giving further confidence that detection is possible and appropiate.

Thermals are found as coherent structures of warm and rising air, surrounded by cold and descending air. A statistical profile is made of the shape of thermals as they appear in timeseries, which is confirmed by LES results. Behaviour of the thermals involving size, velocity and temperature are investigated for different heights.

The average flux within thermals is found to be well modelled by the mass flux approach when based on a thermal surface fraction of 10 percent.

Contents

Chapter 1

Introduction

Vertical transport in the atmospheric boundary layer is usually measured using Eddy-Correlation (EC) techniques. This technique is based on flux averaging in intervals of typically ten or thirty minutes. In reality, however, transport is not limited to these time intervals and will also happen on longer timescales.

In Cabauw, EC measurements are done on four heights, ranging from three to 180 meters. These measurements are used in multiple applications. It is therefore important to quantify the amount of flux represented by longer timescales. This will allow, for example, comparison between the four installations.

The purpose of this paper is to quantify which timescales are important in vertical transport in the atmospheric boundary layer, and which factors play a role in influencing these timescales. The results will be used to find an applicable scheme for predicting the amount of flux found in long timescales.

Also, the multiresolution method as proposed by Vickers and Mahrt (2002) will be investigated to see whether a certain maximum timescale in which turbulent transport happens *(gap timescale)* exists.

Last, the large thermal structures present in the convective boundary layer will be given attention, as they are a large contributor on large scale transport. An attempt will be made to identify these structures from the measured Cabauw time series by comparing sensor data of multiple heights.

The characteristics of these structures will be investigated in an effort combining time series with LES fields. Typical shape and behaviour as conditions vary are studied, as well as the amount of heat flux found in the thermals. The mass flux approach is also evaluated as a method to approximate fluxes within thermals.

Chapter 2

Data and methods

2.1 Data

2.1.1 Cabauw data

For this report, data is used from the Cabauw site. This site features a 213 m high tower equiped with instruments of different sorts to continually monitor the atmospheric boundary layer. Most extensively used in this paper are the Gill R3 sonic anemometers which are mounted on the tower, positioned at respective heights of 3 m, 60 m, 100 m and 180 m.

These produce 10 Hz measurements of the threedimensional wind vector and of temperature, which allows the calculation of Eddy Correlation fluxes.

Since we are interested in longer timescales, a single day will not be sufficient to base our analysis on. We therefore looked for comparable days.

To start, data is used from a rare happening of nine successive cloudless days from May 3rd up to and including May 11th, 2008, which show very nice convection-driven behaviour. This set is then extended with comparable days throughout spring and summer of 2009. This yields a set of seventeen days with temperatures typically between 15 and 25 degrees Celcius, wind speeds below 10 m/s and a wind direction between north- and southeast.

In order to provide a stronger statistical basis for testing of parameterisation or for analysing parameters which are extremely noise, the dataset is further extended with convective days in the summer and spring of 2007 and 2006. These days share the same characteristics, although increasing the temperature spread to a maximum of 32 ^{o}C. This adds another 15 days, giving a total of 32 days with excellent convective data on four measurement heights at high frequency. Table 2.1.1 gives an overview.

2.1.2 LES data

To get a firmer grip on the three-dimensional dynamics of the boundary layer, especially when investigating thermal structures, the measured Cabauw data set is supplemented with LES (Large Eddy Simulation) results. The notion of LES is described extensively in the literature and it is assumed the reader is familiar with it.

An LES run is done with a surface flux of 60 W/m^2 and geostrophic wind forcing of 4 m/s,

Table 2.1: Typical wind conditions of the used data

Typical u [m/s]	Wind direction	Dates
< 4	NE	-
< 4	E	2007/7/17-19, 2007/4/15
< 4	SE	2008/5/7
$4 - 5$	NE	-
$4 - 5$	E	2008/5/10, 2009/8/6, 2006/7/1, 2006/7/16
$4 - 5$	SE	2009/8/5
$5 - 6$	NE	2009/4/29
$5 - 6$	E	2008/5/8, 2006/7/2-3
$5 - 6$	SE	2008/5/3-4, 2008/5/11, 2009/8/4, 2009/8/23, 2006/6/10
$6 - 7$	NE	2009/4/30, 2007/4/2
$6 - 7$	E	2008/5/5-6, 2007/4/30, 2007/5/1
$6 - 7$	SE	-
≥ 7	NE	2008/4/21, 2007/4/1, 2007/4/29
≥ 7	E	2009/4/15
≥ 7	SE	2008/5/9, 2009/4/24

to accomodate fair comparison with Cabauw measurements. A full horizontal cross section is stored at 6 different heights every two seconds to be able to analyse the behaviour of large structures in time. This data is stored at several one-hour periods such that a range of inversion heights is acquired.

The LES experiments are performed without any moisture content, such that virtual temperatures are equal to the absolute temperatures.

Other LES simulations involve one with geostrophic wind forcing of 2 m/s and two simulations with surface fluxes of 30 and 90 W/m^2.

Runs are made with a resolution of 40m horizontally and 20m vertically, on a 128 by 128 horizontal grid, thus spanning more than 5x5km.

2.2 Wavelet Transform

One of the methods that will be used to analyse the given data is wavelet analysis. Wavelet analysis has the great advantage over Fourier analysis that the result is localized in time and space, such that structures in space are not lost, while it still decomposes the series over a range of timescales.

In this report a simple form of the wavelet analysis will be done, as described by de Haij, 2005. Therefore, here we give only a short summary.

The wavelet is a zero-mean function which is localized in time and space, and preferably represents the form of the structures one is looking for. Therefore, the Mexican Hat wavelet is chosen in this research, which is visualized in figure 2.1. It is given by:

$$\psi(x) = \frac{1 - x^2}{\sqrt{2\pi}} e^{-x^2/2} \tag{2.1}$$

Where x is a dimensionless parameter, which represents t/s in our analysis, with t the time
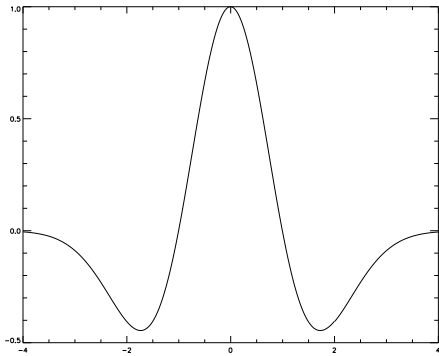
Figure 2.1: General shape of the Mexican Hat wavelet. The function shown here is not normalized.

difference between two points in the time series, and s the wavelet scale. Therefore, the Mexican Hat has zeros at $t = -s$ and $t = s$. The wavelet transform is the convolution of the wavelet function with the time series, so applying a filter which tests the power found in the time series in timescales of the order s.

This transform is then done with increasing s, which can be increased exponentially to avoid using an unnecessary amount of computer resources:

$$s_j = s_0 2^{j/2} \tag{2.2}$$

The smallest scale is chosen as $s_0 = 2dt$. The result of this process is a wavelet transform $W_i(s)$ which is dependent on time (through the discrete timesteps i) and on scale s. To analyse how much power can be found on each scale the function can be averaged over time such that a power spectrum is yielded as follows.

$$\overline{W}^2(s) = \frac{1}{N} \sum_{i=0}^{N-1} |W_i(s)|^2 \tag{2.3}$$

This could be used to find on which scales the greatest amount of power is located by locating the maximum of the above function. Note however that this scale represents the distance from the maximum to the zero crossing, and hence the size is twice this scale.

Chapter 3

Timescales in the atmospheric boundary layer

3.1 Defining a dominant timescale

An important aspect of this study will be to analyze the timescales involved in turbulent transport in the boundary layer. The dominant timescale is relevant in determining the appropiate averaging period when applying the Eddy Correlation method. It also gives - together with the mean wind speed - information on the mean size of turbulent eddies. It is therefore likely that a possible presence of so-called Turbulent Organized Structures (TOS) can also be related to the timescale of transport.

In this paper quantitative relations between these phenomena are sought.

The first problem in this is the determination of the dominant timescale. Many timescales have been proposed, based on, for example, the use of Ogives, the integration of the correlation function (Ouwersloot, 2009), or the integration of the spectrum (Schalkwijk, 2008).

In this paper, the timescales will be based on the cospectra of $\overline{w'T'}$, since it is the dominant timescale in turbulent transport of heat that is the main focus of this research. As a spectrum decomposes time series over frequency, it is a logical candidate to examine when evaluating timescales. The cospectrum will be used as a weight function to determine a weighted frequency. This frequency is then converted to the dominant timescale as $f^* = 1/t^*$.

$$t^* = \left[\frac{\int_{-\infty}^{\infty} f S_{wT}(f) df}{\int_{-\infty}^{\infty} S_{wT}(f) df} \right]^{-1} \tag{3.1}$$

Note that, assuming a certain common spectrum shape, most forms of a dominant timescale can be directly related to each other, and hence the main effect of a choice of a certain timescale involves the way it has to be calculated and the possible errors in this. Some of these errors have been discussed by Ouwersloot (2009).

In the following we rely on surface-layer similarity theory which states that in the lowest part of the atmospheric boundary layer the relevant parameters are friction, velocity and virtual temperature flux. From this, only one dimensionless stability parameter can be constructed, e.g. z/L. Dimensional arguments then show that each dimensionless turbulent quantity can only be a function of z/L.

For observations higher up in the boundary layer, this approximation will fail and and other parameters, like boundary layer height, come into play.

3.2 Spectral analysis

As suggested in the previous section, spectral decomposition is very useful when examining timescales in the boundary layer. Kaimal (1972) showed that spectra demonstrate great similarity when made dimensionless with friction velocity u^* and temperature scale T^* and grouped by the stability parameter z/L. In this paper a somewhat simplified approach will be taken, where the surface layer spectra approximation is used according to Bosveld, 1999, which is based on Kaimal's paper.

The relation is as follows, where n is the dimensionless frequency $n = fz/U$.

$$\frac{U}{z}S^f_{wT}(f) = S^n_{wT}(n) = F_T\frac{a}{\left(1 + n\frac{a}{p-1}\right)^p}$$ (3.2)

Which is here normalized such that $\int_0^\infty S^n_{wc}(n)dn = F_c$. Now introducing the earlier proposed dominant frequency f^*:

$$f^* = \frac{1}{F_T}\int_0^\infty fS^f_{wT}(f)df = \frac{U}{zFc}\int_0^\infty nS^n_{wT}(n)dn$$ (3.3)

Introducing 3.2, this results in:

$$f^* = \frac{U}{z}\int_0^\infty \frac{an}{\left(1 + n\frac{a}{p-1}\right)^p}dn$$ (3.4)

Now changing variables to $x = 1 + \frac{an}{p-1}$, the integral becomes:

$$f^* = \frac{U}{az}\int_1^\infty \frac{(x-1)(p-1)^2}{x^p}dx$$ (3.5)

The integral can now be performed analytically for $p > 2$, which is the relevant case for the opted application, resulting in:

$$f^* = \frac{U(1-p)}{za(2-p)}$$ (3.6)

In the case of the turbulent heat fluxes: $w'T'$ we have $p = 7/3$ and hence we find:

$$a = \frac{4U}{f^*z}$$ (3.7)

Note that this equation allows us to compare the dominant timescale $t^* = 1/f^*$ with the timescale used in Bosveld's analysis, J_{wc}. Bosveld defined the timescale as:

$$J_{wc} = \frac{1}{R_{wc}(0)}\int_{-\infty}^\infty R_{wc}(\tau)d\tau$$ (3.8)

Which resulted in $a = 2J_{wc}U/z$. Comparing this with 3.7 we find the direct relationship

$$t^* = \frac{1}{2}J_{wt} \tag{3.9}$$

Returning to the spectrum, inserting a and setting $n^* = f^*z/U$, we find:

$$S_{wT}^n(n) = \frac{4F_T}{n^*}\frac{1}{\left(1 + \frac{4n/n^*}{p-1}\right)^p} \tag{3.10}$$

Interesting is the fact that we can rewrite this spectrum in terms of f instead of n to find:

$$S_{wT}^f(f) = \frac{4F_T t^*}{\left(1 + \frac{4ft^*}{p-1}\right)^p} \tag{3.11}$$

This allows some further physical interpretation of the earlier defined timescale t^*. Note that, when first dividing equation 3.11 by $F_c = \int_0^\infty S_{wc}(f)df$ and then multiplying left and right with f, the right hand side becomes a function of f/f^* only. This allows averaging of co-spectra derived from different periods, which do not need to have the same timescale t^*. Since t^* is also dependent on z/L this means that such an average may also contain co-spectra with different stability. This is an obvious advantage over displaying the co-spectra as a function of the dimensionless frequency n.

Averaging in this paper is done by dividing the horizontal scale (in this case ft^*) in a suitable number of bins. The data is then grouped into these bins. The data within the bins is averaged. Horizontal bars over the points display the bin size, vertical bars the *error in the mean* . Hence the vertical error bars represent the error of a certain point $\overline{y}$ given by:

$$\sigma_{\overline{y}} = \frac{\sigma_y}{\sqrt{N}} = \frac{1}{\sqrt{N}\sqrt{N-1}}\sqrt{\sum_{j=0}^{N-1}(y_j - \overline{y})^2} \tag{3.12}$$

Where N is the number of points y in the given bin, and σ_y the standard deviation of y. Important to note is that while a small vertical error bar indicates a small error in the mean, no error bar is drawn in the bins where $N = 1$, meaning an unkown error instead of no error.

Figure 3.1 shows such an averaged spectrum. It shows a comparison between measured spectral data and equation 3.11, based on surface layer theory. The data is taken from 10-Hz sonic anemometer observations, collected over seventeen selected convective days, spread over spring and summer of various years. The data is corrected for tilt before spectral analysis is done. Cross-wind correction on sonic temperature is done internally in the instrument. Spectra are based on three-hour intervals around noon. These intervals are split into sections of 6 minutes for 3m, and 14 minutes for other heights. These section lengths are chosen as it has been found that the timescale t^* is independent of further increasing the maximum section length, suggesting that these section lengths are long enough for a correct evaluation of t^*.

Figure 3.1 shows the spectra on a logarithmic horizontal axis. The result is that a peak is introduced in the spectrum. This peak is intuitively a measure of which timescales are important in $\overline{w'T'}$ transport. Indeed, using equation 3.11 by setting the derivative of $fS_{wc}(f)$
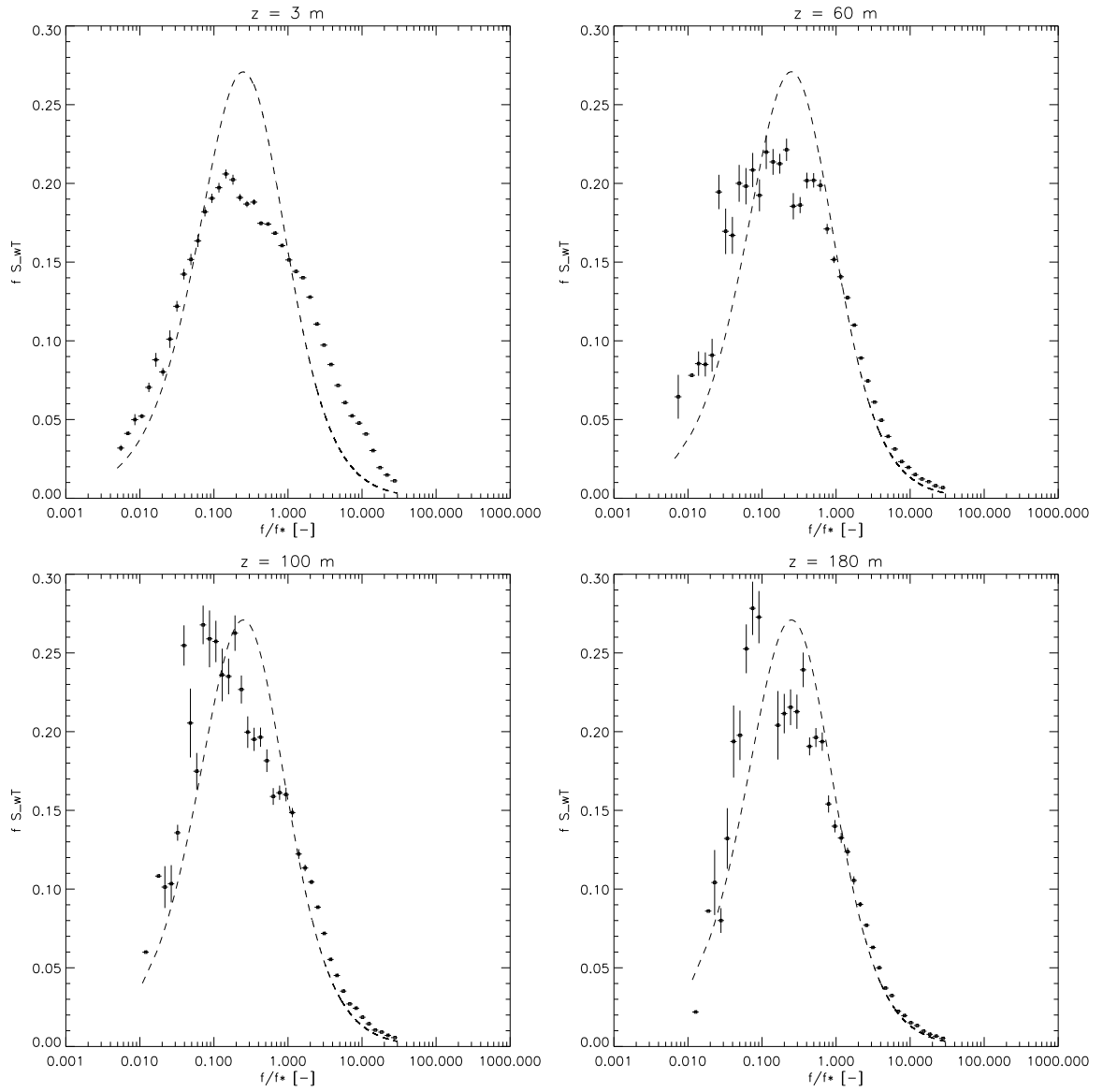
Figure 3.1: Spectral data of a three-hour interval around noon (dots), averaged over 32 convective days, is compared to the empirical form of equation 3.11 (dashed line).

to zero, it is found that the peak is located at a frequency corresponding to $t = 4t^*$. This is confirmed when examining figure 3.1.

The qualitative behaviour of the observed co-spectra is reasonably modelled by the simple relationship of equation 3.11, especially on the high-frequency range. This in accordance with, for example, Kaimal, 1972, who finds noisy behaviour in the low frequency part of spectra for $z/L < 0$, but better behaviour in the inertial subrange. Note that all the data used for figure 3.1 indeed has $z/L < 0$ as can be expected for convective days.

3.3 Low frequency loss

The true flux at a given time is defined as the rate of transport through a given horizontal plane. When measuring Eddy-correlation (EC) fluxes, temporal averages are used to approximate the true flux. In this process, the underlying assumption is that when measuring long enough, a representative image of the flux is formed. A prerequisite for this to be true is thus, that all characteristics of the flux field 'pass by' the observer in this averaging period. However, fluxes which are transported over timescales longer than the averaging time T, will not be adequately represented by the EC-flux. This leads to an overall low frequency loss. In this section an analytical approximation to the low frequency loss is sought.

Let $w(t)$ and $c(t)$ be the stationary turbulent time series of the vertical wind speed and a scalar quantity, respectively. With the overbar denoting temporal averaging, the Eddy Covariance flux over a period T is found as:

$$F_c^T = \frac{1}{T} \int_{-T/2}^{T/2} (w - \overline{w}) (c - \overline{c}) \, dt \tag{3.13}$$

Writing out the product and performing ensemble averaging (denoted with the curved parenthesis) the actual flux F_c can be taken out and, introducing the time correlation function $R_{wc}(\tau)$, this can be rewritten to:

$$\{F_C^T\} = F_c \left[1 - \frac{1}{T} \int_{-T}^{T} d\tau \left(1 - \frac{|\tau|}{T} \right) R_{wc}(\tau) \right] \tag{3.14}$$

Now performing the integration we find that the loss becomes:

$$Loss = \int_0^\infty S_{wc}^f(f) \frac{\sin \pi f T^2}{(\pi f T)^2} \, df \tag{3.15}$$

Where the cospectrum $S_{wc}(f) = \int_{-\infty}^\infty R_{wc}(\tau) e^{2\pi i f \tau} d\tau$ is introduced. The factor with which the cospectrum is multiplied can be thought of as a low-pass filter, which can be approximated by the simpler expression $1 - fT$ to find:

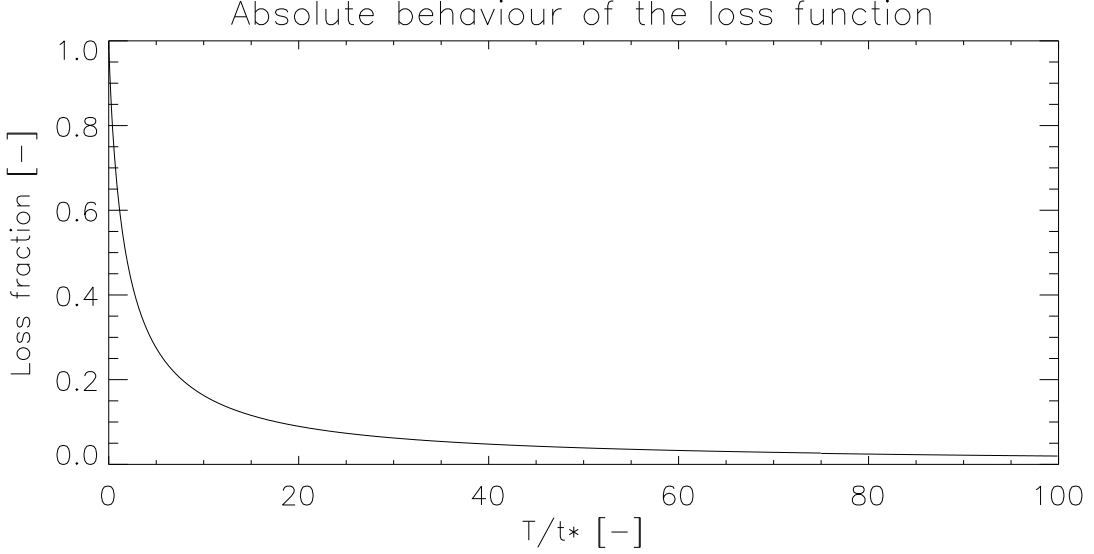$$Loss = \int_0^{1/T} S_{wc}^f(f) (1 - fT) \, df \tag{3.16}$$

Figure 3.2: The absolute behaviour of equation 3.19.

In the previous section an expression was found for the spectral behaviour. Using this relation allows us to find an explicit expression for the loss as follows. Rewrite equation 3.15 for the case of heat fluxes in terms of $S_{wT}^n(n)$ and introduce 3.10 to find:

$$
\begin{aligned}
Loss &= \int_0^{\frac{z}{TU}} S_{wT}^n(n)(1 - \frac{nTU}{z})dn \\
&= \frac{4F_T}{n^*} \int_0^{\frac{z}{TU}} \left(1 + \frac{4n}{n^*(p-1)}\right)^{-p} \left(1 - n\frac{TU}{z}\right) dn
\end{aligned}
\tag{3.17}
$$

Now change variables to $x = 1 + \frac{4n}{n^*(p-1)}$, and introducing $y = 1 + 4/f^*T(p-1)$, we have:

$$
Loss = \frac{F_T(p-1)}{y-1} \int_1^y yx^{-p} - x^{1-p}dx
\tag{3.18}
$$

This integral can be evaluated to yield the final expression:

$$
Loss = F_T \left(1 + \frac{y^{2-p} - 1}{(p-2)(y-1)}\right)
\tag{3.19}
$$

Interesting is that the loss fraction is now solely a function of T/t^*, the averaging time over the dominant transport timescale. Figure 3.2 demonstrates the behaviour of equation 3.19. The loss quickly decreases over the first five t^* to about 30%. It then converges slower to reach a loss of 10% in about $20t^*$ and 5% in $40t^*$.

Now that a physical interpretation of t^* and the behaviour of the Eddy Covariance loss as a function of t^* are known, it is a good idea to evaluate how t^* typically behaves in the atmospheric boundary layer, and look at the typical order of magnitude. Figure 3.3 shows $t^*(z)$ for near-neutral, unstable and stable situations. These situations are chosen so as to

accomodate comparison with the gap timescale which will be introduced in chapter 3.5.

The figure is based on data from 32 days which are especially selected for their optimal convective characteristics. Data with $|L| \geq 100$ is grouped as near-neutral, while data with $|L| < 100$ is grouped by sign into unstable and stable groups. This yields a dataset of respectively 4, 24 and 36 three-hour periods at all measuring heights.

Note that dependence of t^* on other parameters like horizontal wind speed U is expected, so figure 3.3 should be seen as an indication of the typical order of scales in the boundary layer for different situations, and not be interpreted too quantitatively.

The figure shows a behaviour of t^* like what one would expect from a timescale which characterizes turbulent transport. It increases with height, yet the increase becomes less steep with greater height. Note that a slope corresponding to the diagonal of the graph corresponds to a scale proportional to the height. The increase is also influenced by the stability, as in unstable situations the timescale can grow to more than an order larger than in stable situations.

On a clear, convective day, which is a typical unstable situation, the timescales grow from the order of 4 seconds at 3m to more than a minute at 180m. Comparing this with the behaviour of equation 3.19, the result is that averaging times in the order of 20 minutes will be needed at 180m to reduce the loss below 10%. At 100m this drops to 10 minutes, while at 3m an averaging time of 2 minutes is more than enough to reach this accuracy. In stable situations timescales rarely grow beyond the order of 10 seconds, even at greater heights.

The length scale characterizes the typical order of the size of eddies dominantly responsible for turbulent transport. Note that at 180m in convective surroundings, the length scale typically grows to as much as 300m. Remembering that the peak of the $fS_{wc}(f)$ spectrum falls on $t = 4t^*$, this would correspond to eddy sizes on the order of the inversion height, which is what is generally expected to happen in this situation.

Interesting is also to see that length scales are of the same order in unstable and stable situations for measurements 3m high. This is a strong indication that eddy transport at this height is mainly limited by the height itself and independent of other parameters. The differences in time scales in this regions is then solely explained by differences in wind speed.

Returning to the loss function, note that in the process of deriving equation 3.19, no additional assumptions where made in comparison to the result (eq 6.12) in Bosveld, 1999. This means that 3.19 is simply another manifestation of the same function, now based on the earlier defined dominant timescale t^*.

This provides the possibility of comparing these expressions. The advantage of using equation 3.19 is that the t^* is acquired directly from the spectral data. This means that no empirical relations are needed to find the loss, one can rely on measurements only. This provides a nice way of examining the correction method.

First, it is now possible to compare the empirical relation which is used to determine J_{wT} using t^*, since equation 3.9 relates t^* directly to J_{wT}. The advantage in this lies in the fact that t^* is easier to determine since it does not require the integration of an infinite correlation function.

Bosveld parameterizes the timescale for turbulent transport of heat as:

$$l_{wt}(z/L) = \frac{J_{wT}U}{z} = \begin{cases} 5.0 & z/L < 0, \\ \frac{5.0}{1+6.7z/L} & z/L \geq 0. \end{cases} \tag{3.20}$$
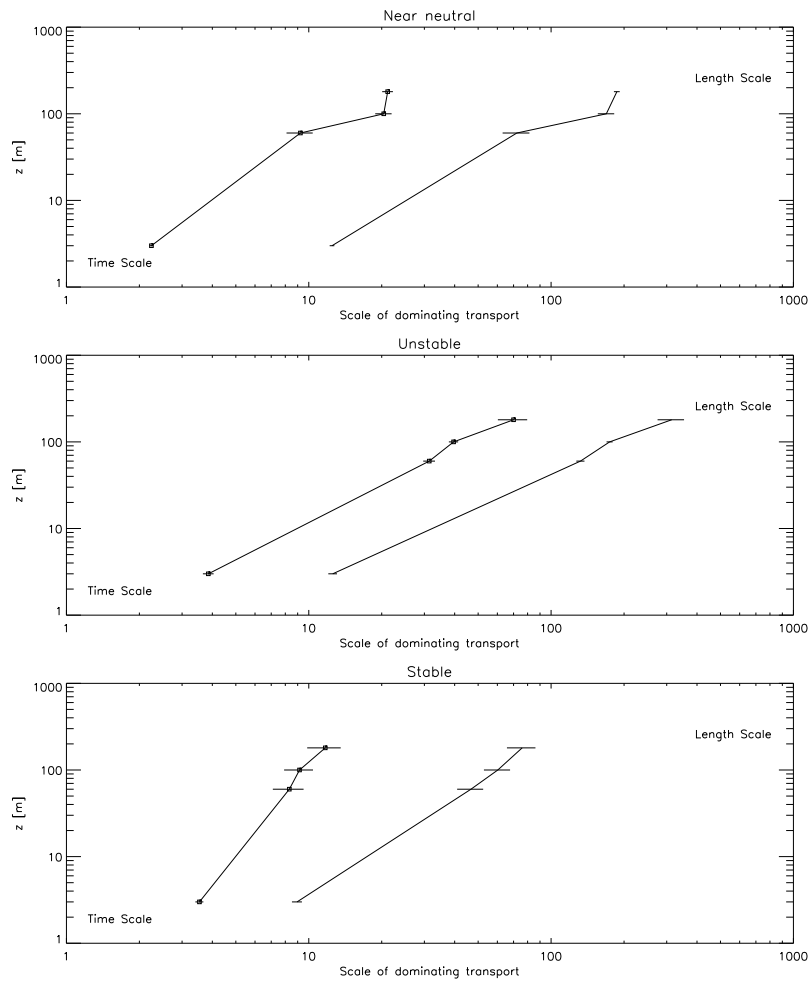
Figure 3.3: The average behaviour of the dominating transport timescale t^* in seconds (connected boxes on the left) and the associated length scale (m) which is found by Taylor's hypothesis as $L^* = Ut^*$ (right line).
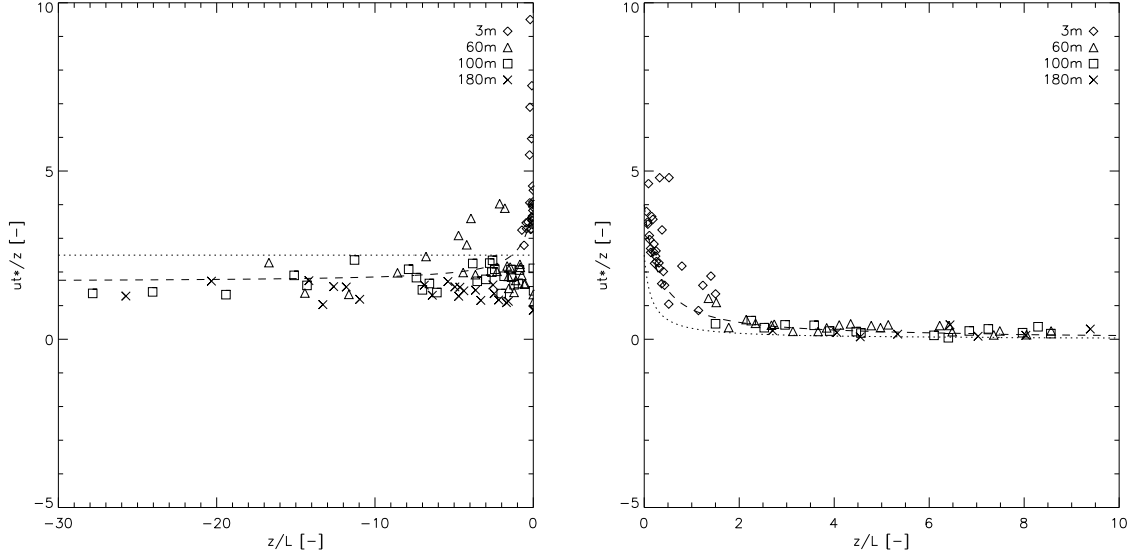
Figure 3.4: The nondimensionalized dominant timescale is plotted against the stability parameter z/L. Data collected at different heights is distinguished by symbol. The dotted line represents the equation 3.21. The dashed line represents a custom fit, equation 3.22. The left figure shows unstable cases for convective circumstances, the right figure shows stable cases at night. Both figures are based on a selection of 32 days.

Using equation 3.9, we can relate this equation to t^*:

$$\frac{Ut^*}{z} = \begin{cases} 2.5 & z/L < 0, \\ \frac{2.5}{1+6.7z/L} & z/L \geq 0. \end{cases} \tag{3.21}$$

Note that this means that for ideal days with very convective circumstances, L is typically negative and hence the parameterisation is simply $Ut^*/z = 2.5$. To evaluate the reliability of this approximation, data is collected from 32 days, selected for their convective circumstances. For each of these days, two periods are selected, one from 0.30 to 3.30h at night, and one from 10.30 to 13.30h around noon. Measured data is corrected for tilt and then averaged over these periods. The results are shown in figure 3.4.

The dotted line in figure 3.4 represents equation 3.21. The figure suggests that in convective circumstances, data from the 3m sensor is structurally being underestimated by equation 3.21, data from the higher sensors is overestimated. In stable circumstances, all data is underestimated.

The dashed line in figure 3.4 is a custom fit:

$$\frac{Ut^*}{z} = \begin{cases} 1.70 + \frac{2.30}{1-1.43z/L} & z/L < 0, \\ \frac{4.0}{1+3.5z/L} & z/L \geq 0. \end{cases} \tag{3.22}$$

Which is seen to better match the data than equation 3.21. Note that closer inspection suggests that influence from an extra parameter, as data from increasing heights seems to structurally lie slightly lower in the graph. This unkown parameter is expected to be the boundary layer height z_i. Indeed, a slight structural dependence on z_i is found, but for the

intended application the slight induced error by using a single curve for all heights is not very significant, and the fitted curve is clearly an improvement on equation 3.21.

The problem in equation 3.21 lies in the way it is derived. Note that, in the process of acquiring equation 3.11, the spectrum is modelled by a simplified curve. The peak of this curve lies in the point $n = 1/(2l_{wc})$. Equation 3.20 is based on this relation and the behaviour of the maximum, as is found in Kaimal, 1972. The problem here is that equation 3.11 is based on the $-7/3$ power spectrum, while Kaimal uses a function composed of two parts, where the $-7/3$ powered part is typically *not* the part where the maximum lies.

For example, Kaimal introduces the following parameterisation for the case $z/L = 0$:

$$nS_{wc}^n(n) \sim \begin{cases} \frac{11n}{(1+13.3n)^{1.75}} & n \leq 1, \\ \frac{4.4n}{(1+3.8n)^{2.4}} & n > 1. \end{cases} \tag{3.23}$$

Differentiating, the maximum of $nS_{wc}(n)$ lies in the low frequency part of the function, at $n = 0.10$. This directly results in the 5.0 to which equation 3.20 reduces when $z/L = 0$. However, it is clear that the $-7/3$ part of the function is no longer valid here.

How much of a problem this assumption exactly is can only become clear by looking at the correction scheme at work. Using raw 10Hz data from the Cabauw tower, it is possible to study the effects of varying the averaging period. This can then be compared with the relationships derived above.

Figure 3.5 does exactly that. The diamonds represent the $\overline{w'T'}$ flux as the averaging time varies. The data is based on 3-hour intervals around noon on the selected days. Only averaging times which fit an integer number of times into the 3-hour interval - such that no data is missed - are used. The data is then normalized as relative loss with respect to the 5 or 20 minutes averaged flux, for 3 meter or other heights, respectively:

$$L = 1 - \frac{\overline{w'T'}(T_{averaging})}{\overline{w'T'}(T_{max})} \tag{3.24}$$

All data has been corrected for tilt before it is used.

The solid line represents equation 3.19, also normalized with respect to the same averaging time. This allows comparison of the behaviour of the fluxes. Note that, as t^* comes from the spectral data, the only approximations introduced in this line are the use of the spectral shape as proposed by Kaimal, equation 3.2 and the simplification of the low pass filter, equation 3.16. The dots represents how the correction algorithm is now applied, hence it introduces another approximation, equation 3.21. Calculated values of t^* and $J/2$ are also displayed. When these two equal, the dots fall exactly over the solid line.

The interpretation of this figure is now that points which fall above the solid line are fluxes with smaller magnitude than surface layer theory would predict. Where points fall under the solid line, surface layer theory underestimates the true flux.

Figure 3.5 demonstrates that for very convective days, fluxes behave very reasonably according to surface layer theory all the way up to 180m high. The data does become more noisy at greater altitude, but the loss is estimated reasonably well by the correction algorithm, even though t^* is not always accurately represented through $J/2$.

Although figure 3.5 illustrates well the principal idea of the correction method, it does not provide enough statistics to give much information on the quantitative performance of the
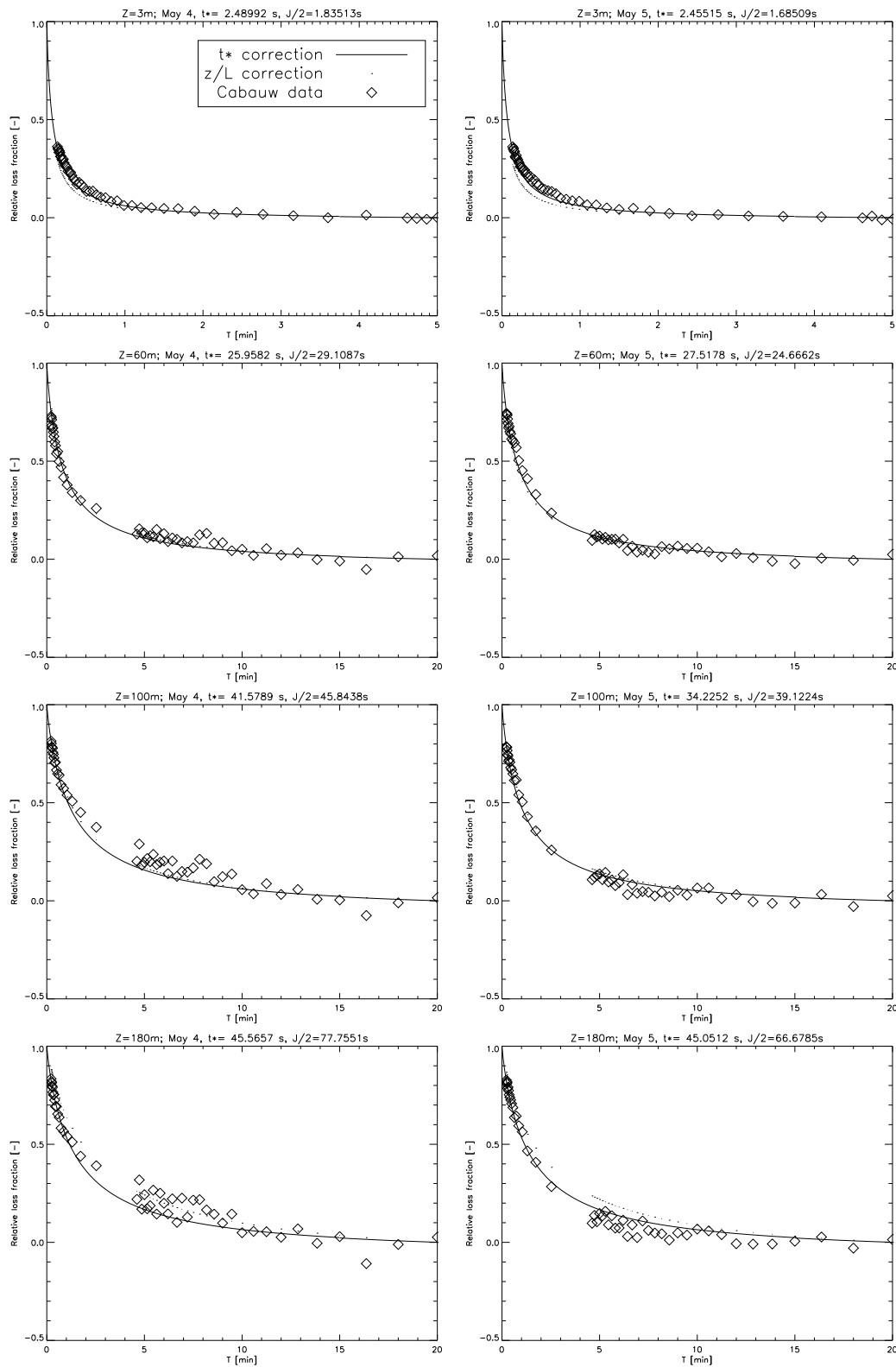
Figure 3.5: The behaviour of $\overline{w'T'}$ fluxes as averaging time varies is compared with correction algorithms as derived from boundary layer theory. The diamonds represent data taken from the Cabauw tower, measured on May 4 and May 5, 2008. The solid line is equation 3.19 with t^* from the spectral data, while the dots(.) result from the use of relationship 3.21.

method. To better evaluate whether or not systematic errors are made, the averaging algorithm discussed earlier is used to average the behaviour of the measured flux as a function of T/t^*, the averaging time over the timescale. Note that equation 3.19 suggests that indeed the loss is a function of this quantity only.

Figure 3.6 shows this averaged behaviour, with the black boxes representing the averaged data and the red crosses representing the averaged prediction of the correction algorithm, based on the same data. The only difference between the red crosses and the dotted line lies in the way t^* is found, which the red crosses derive from equation 3.21 while the dotted line is based on the spectrum. Hence, the impact of the error visible in figure 3.4 is directly visualised.

In this figure, none of the data is normalized such that it accounts to zero after a certain period of time (equation 3.24). Instead, $\overline{w'T'}(T_{max})$ in equation 3.24 is corrected according to equation 3.19. This forces $\overline{w'T'}(T_{max})$ on the dashed line. However, this action is less drastic than forcing the loss to be zero at T_{max} in order to enable comparison, which is what is done if the loss is displayed relative to the flux at T_{max}. Therefore, figure 3.6 allows for some more quantitative evaluation.

The general picture of figure 3.6 is rather promising. Especially at 60 to 100 meters high the correction algorithm (red) follows theory very well. This could have been expected from figure 3.4, which shows that many of the 60 and 100 meter measurements are found at $z/L \sim -5$ and scatter close to the dashed line. Even more promising is the fact that the data from the Cabauw tower falls very nice over the correction predictions, indicating that the behaviour of the fluxes when varying the EC averaging time can indeed be predicted.

At 3m and 180m, figure 3.6 shows a departure by the correction method from theory. This can also be seen from figure 3.4, which shows that the value of Ut^*/z is systematically overestimated at 3m and systematically underestimated at 180m. The result is that the correction predicts a faster respectively slower decrease to zero loss. Since the data still closely follows the predicted loss (dashed line), an error is introduced in the correction.

For average use, the error in the correction at 3m is not a large problem, as it has largely diminished after $T/t^* = 50$, which at this height typically occurs at an averaging time of 2.5 minutes, where typical averaging times stretch over 10 minutes or more. At 180m, however, averaging times of 10 minutes typically correspond to $T/t^* = 10$, at which point a significant departure of the correction method is found.

The blue triangles represent the result of using Bosveld's correction method, but now based on the fit of equation 3.22. As expected, it is seen to match theory and data better for all heights, but especially for 3m and 180m. Although for $T/t^* > 10$ at 180m the performance of the correction decreases, it remains better than or equivalent to the performance of equation 3.21.

For convective days then, it can be concluded that surface layer theory holds well enough and the correction algorithm based on these relationships can well be used. When tuning the correction method with equation 3.22 the correction method's applicability can be extended up to measurements at 180m height.

Figure 3.7 shows the behaviour in stable situations. The same 32 days are used, but now a period from 0.30h to 3.30h is selected. Cases with an absolute flux $|\overline{w'T'}| < 0.010 Km/s$ are left out since this introduces practically only noise. The result of this is that only heights of 3m and 60m are displayed since above these heights the absolute flux rarely meets this
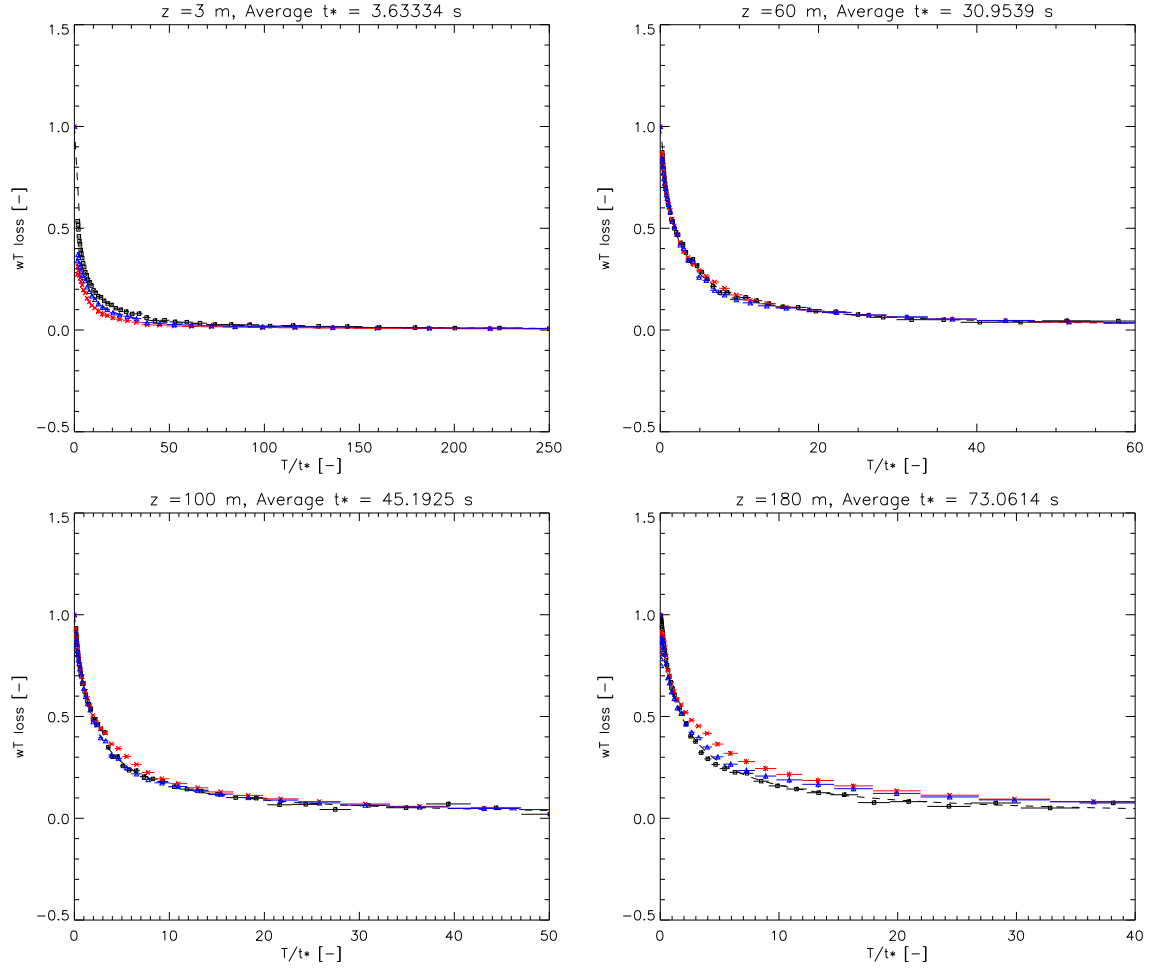
Figure 3.6: The behaviour of $\overline{w'T'}$ fluxes as a averaging time varies is compared with correction algorithms as derived from surface layer theory. Black boxes represent data which is averaged over 32 days which are selected for their convective circumstances, with the average value for t^* printed in the title. Averaging bin sizes vary exponentially such that like amounts of points fall in each bin. The red crosses represent the averages of the correction method described in Bosveld, 1999, applied on the same dataset. The dotted line represents equation 3.19 with t^* extra directly from the spectra. Blue triangles represent the averages of a correction method equal to Bosveld, but based on equation 3.22
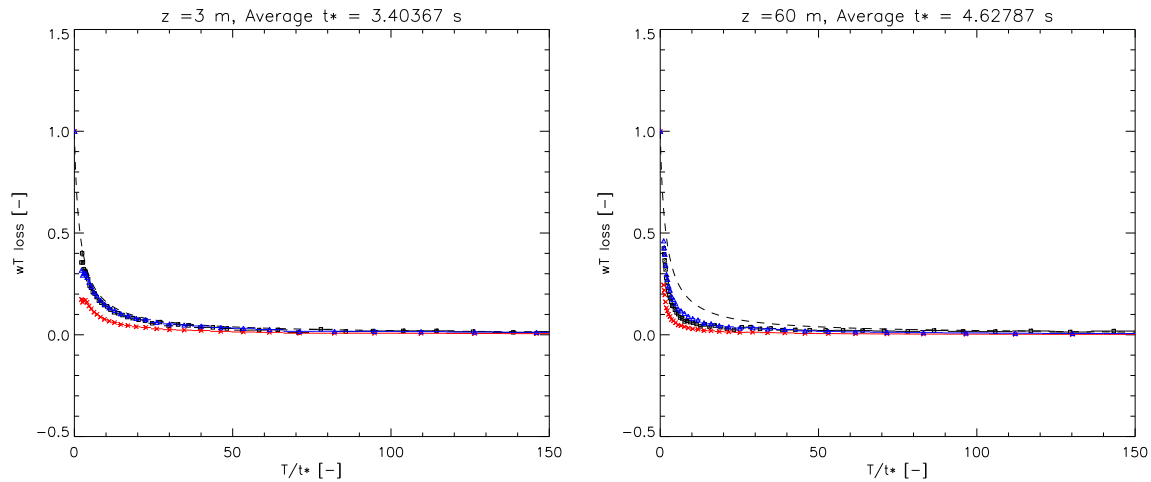
Figure 3.7: The behaviour of $\overline{w'T'}$ fluxes as a averaging time varies is compared with correction algorithms as derived from surface layer theory. Black boxes represent data which is averaged over 32 days. Red crosses represent the averages of the correction method described in Bosveld, 1999, applied on the same dataset. The dotted line represents equation 3.19 with t^* extra directly from the spectra. Blue triangles represent the averages of a correction method equal to Bosveld, but based on equation 3.22

minimum criteria.

At night, the correction algorithm (red) systematically underestimates the loss, while a tweak with equation 3.22 (blue) shows a striking improvement. At 3m high, the figure shows great agreement between theory, use of equation 3.22 and the data.

The plot shows interesting features at 60m as well. Note that the dashed loss function seems to overestimate the loss systematically, while the blue line for the tuned correction method follows the data almost perfectly. Since the only difference between these methods is the way t^* is evaluated, a possible explanation for this lies in the fact that the spectra of the smaller nightly fluxes are typically much noisier than the spectra for fluxes in convective surroundings. Hence, it is possible that evaluating t^* using this spectra (equation 3.1) is less succesful at night.

While these differences are striking for $T/t^* < 20$, they have little impact for the typical averaging time of 10 minutes, since for the small values of t^* found at night, this would correspond to values of T/t^* around 150, at which point the fluxes have long converged.

The major sidenote on figures 3.6 and 3.7 would be that correcting $\overline{w'T'}(T_{max})$ according equation 3.19 forces the data at T_{max} to fall on the dotted line. Although the agreement in the behaviour of the data and predictions is very good, which at least indicates the correctness of the prediction, it gives no information on the fluxes found on timescales larger than T_{max}, and hence the absolute loss found in the data.

To investigate whether or not a significant flux is found in timescales larger than evaluated above, and which are not predicted by the above correction methods, the above analysis is continued for longer timescales. T_{max} is taken at the same values of T/t^* as was done for figure 3.6, but after $T = T_{max}$ the analysis is continued. This way the data is allowed to deviate from the prediction for $T > T_{max}$ if a significant flux is found there.
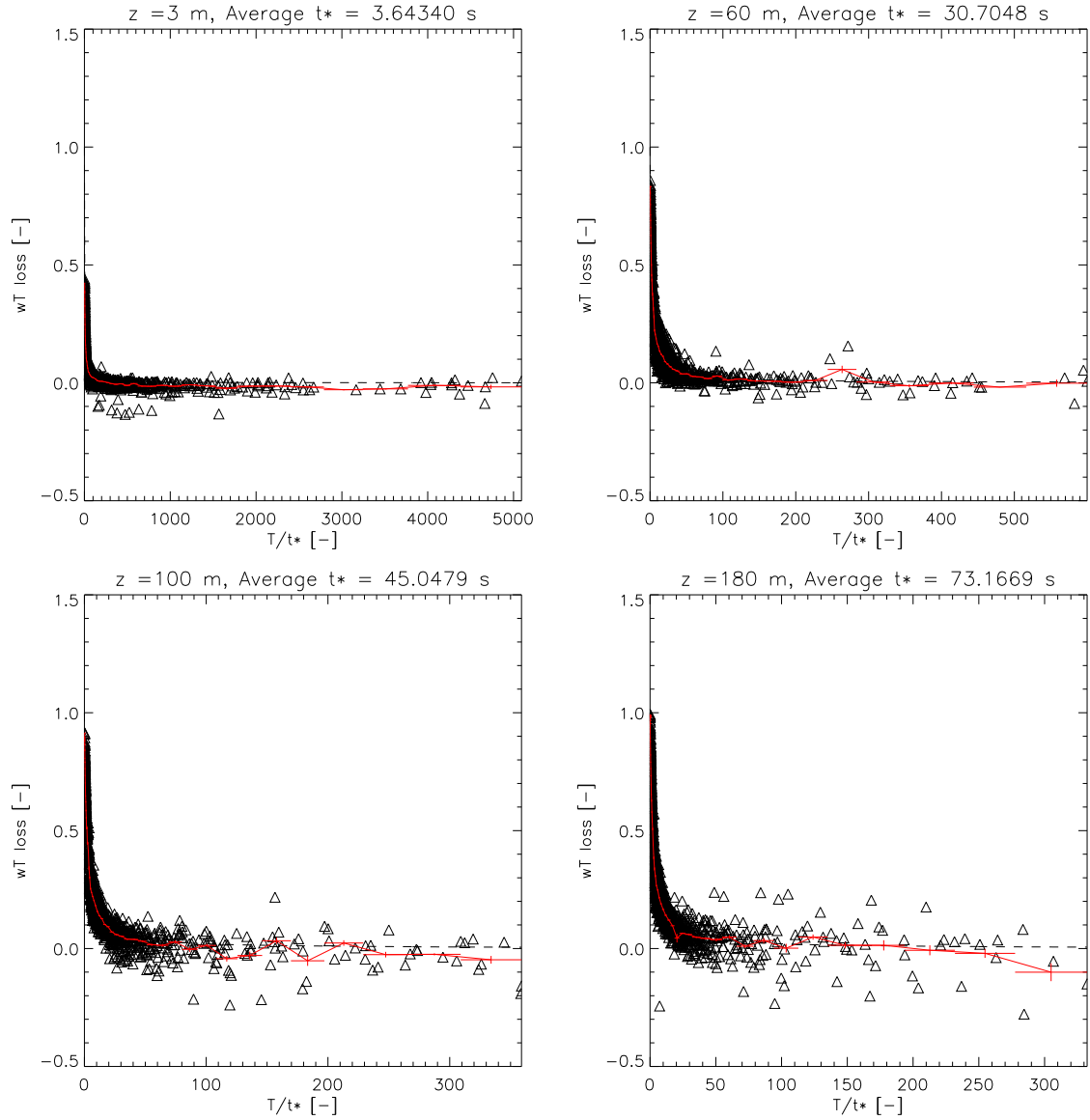
Figure 3.8: The behaviour of $\overline{w'T'}$ fluxes as a averaging time varies is compared with correction algorithms as derived from surface layer theory. Black triangles represent individual results, while the red line pluses display the result of averaging the individual results over the bin size indicated by the horizontal line. The dashed line represent the theoretic loss of equation 3.19.

Figure 3.8 displays these results. Note that the individual results show an increased spread as T/t^* increases, which also increases with height. One explanation for this spread is the fact that points with higher T/t^* fit less integer times into the interval used for calculation, increasing the uncertainty. Another might be the influence of diurnal variability. An attempt is made to compensate for this increase of uncertainty by averaging over a large amount of days and evaluating if an average systematic flux is found. The results of this averaging is the red line. Figure 3.8 shows data based on three-hour intervals around noon of 32 selected days. For this figure, used temperature data is first detrended. Note that this involves the possibility of removing a part of very long timescale fluxes, but it is found that the spread is greatly decreased by detrending. This suggests the detrending process did not remove any systematic flux.

Figure 3.8 gives no reason to suggest a significant, systematic flux at larger timescales for the lower heights. At 3m and 60m the flux is very well-behaved. However, the figure shows an increasing mismatch with the predicted loss for increasing heights. For 100m and 180m, it seems that an underestimation of the true flux is made by small-timescale measurements, since the spread of points seems to be centered beneath the dashed line. However, the amount of datapoints is not large enough to provide hard proof on this subject. Even besides the amount of datapoints, it is the question if errors in eddy covariance averaging can be found using EC methods, even when averaging for long timescales. After all, if the EC method converges to the wrong value (as has been suggested by Schalkwijk, 2008; and Ouwersloot, 2009), this will never be detected using above methods.

The issue here is that with timeseries at one point we will never sample all relevant structures occuring at a large horizontal plane.

It seems techniques as used so far cannot be used to prove whether a systematic part of the flux is found in very large timescales. Note that comparison is further made difficult by the fact that these papers have been mainly concerned by the fundamentals of such a systematic loss, and have only studied cases for $U = 0$. The loss correction as described in this paper is a function of T/t^* (equation 3.19), where t^* is found by equation 3.22. Note that this last equation behaves asymptotically for t^* as U approaches zero, hence the loss prediction is undefined for this case.

3.4   Eddy Covariance flux measurements in LES data

Later in this report, several qualities of behaviour in timeseries from Cabauw data can be reproduced in LES simulations. This raises the question whether the behaviour of EC flux measurements as a function of timescales can also be reproduced. In fact, LES data has a major advantage over measured data when considering large scale contributions: the 'true' flux is very closely approximated by the spatially averaged flux, and hence we are not dependent on the correction of a long-averaged flux to represent the true flux, as we were in the production of figure 3.6.

Therefore, the true loss can be visualized, while it becomes possible to investigate whether the EC flux really approaches the spatial flux for long timescales.

This advantage comes at a cost, though, since LES simulations at common resolutions cannot resolve turbulence nearly as small as the 10 Hz sonic measurements can. This causes problems

when calculating the dominant timescale t^* as was done for figure 3.6. To compensate for this, t^* is corrected for this high frequency error based on a general spectrum shape (equation 3.11).

For a fair comparison, figure 3.4 shows the behaviour of the loss completely analoguous to what was done for figure 3.6, meaning that the loss is considered with respect to the corrected EC flux at T_{max} and not the true flux. The figure is now however based on 100 randomnly chosen time series from a 128 by 128 horizontal grid at heights approximately corresponding to the top turbulent observation levels of the Cabauw tower. The time series are based on the LES from 9 to 10 hours simulated time, in which the boundary layer height grew to the order of 1km.

The behaviour is very much alike the Cabauw data loss, which is a comforting knowledge. However, the theoretic line seems to systematically underestimate the true loss for larger T/t^*. Since the behaviour for low T/t^* is very good, it is difficult to state whether this is due to a faulty evaluation of t^*. This would be the first expected cause, as it is not unexpected that the correction of t^* is not completely accurate.

Now that the similarity is established, the same analysis can be done, but now regarding the loss with respect to the spatial flux, which is supposed to represent the 'true' flux. Figure 3.10 shows this results. The same qualitative behaviour is witnessed as in figure 3.6 for small T/t^*, although it does not completely correspond to the theoretical line, especially at greater altitude. This might point at a faulty evaluation of t^*.

However, for larger T/t^*, the EC average stops converging to the true flux, resulting in a typical absolute loss of about 20%, with this absolute loss decreasing with height. This is a strange phenomena which can not easily be explained, as expectations are that the loss increases with height (Schalkwijk, 2008; Ouwersloot, 2009). The question is also whether the behaviour for large T/t^* in the first figure 'warned' us of the behaviour for the true loss. This would indicate that as the measurements show no such behaviour, they do not deviate from the true flux that much.

A full LES study will be needed to state whether or not the EC behaviour in LES is fully comparable to the behaviour of measured time series.

3.5 Multiresolution decomposition

Instead of using a Fourier analysis to find a spectrum, Vickers and Mahrt (2002) propose the use of multiresolution spectra. This algorithm decomposes the flux over timescales $t = \delta t 2^m$, where m integer. This is done by partitioning the interval into averages on scales of 2^m consecutive data points. Where Fourier analysis decomposes the spectrum using predefined functions $e^{2\pi i n t/T}$, multiresolution (MR) analysis decomposes the record into averages on different timescales, satisfying Reynold's averaging on all scales. A full description of the MR algorithm is found in Vickers and Mahrt, 2002.

According to Vickers and Mahrt, the MR cospectrum D_{wc} relates to the Fourier spectrum S_{wc} as $D_{wc}(m) \sim f S_{wc}(f)$. This would mean that the MR spectrum can be related to the modelled spectrum 3.11 directly, according to:

$$D_{wc}(m) \sim \frac{4F_C t^*/t}{\left(1 + \frac{4t^*/t}{p-1}\right)^p} \tag{3.25}$$
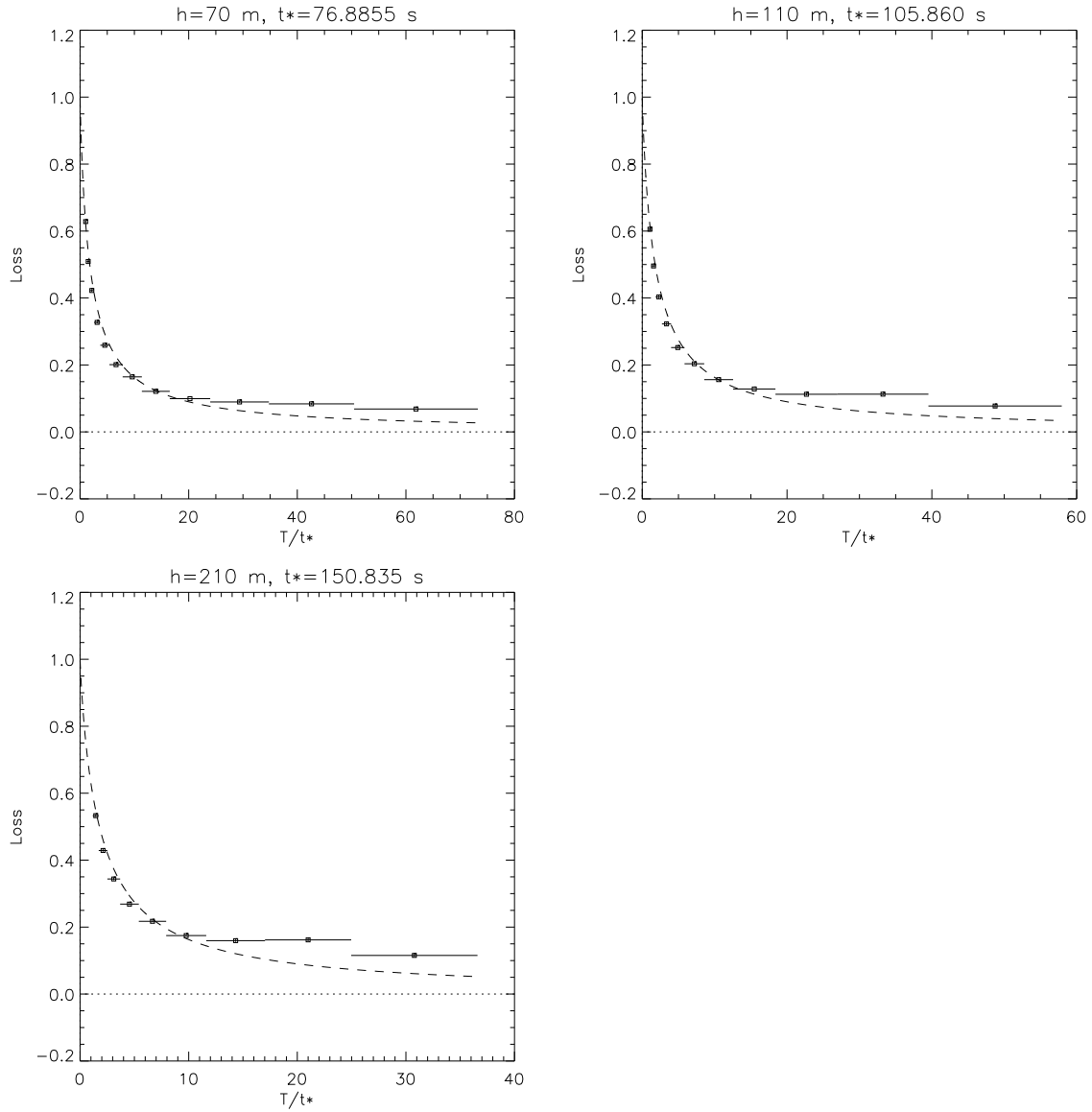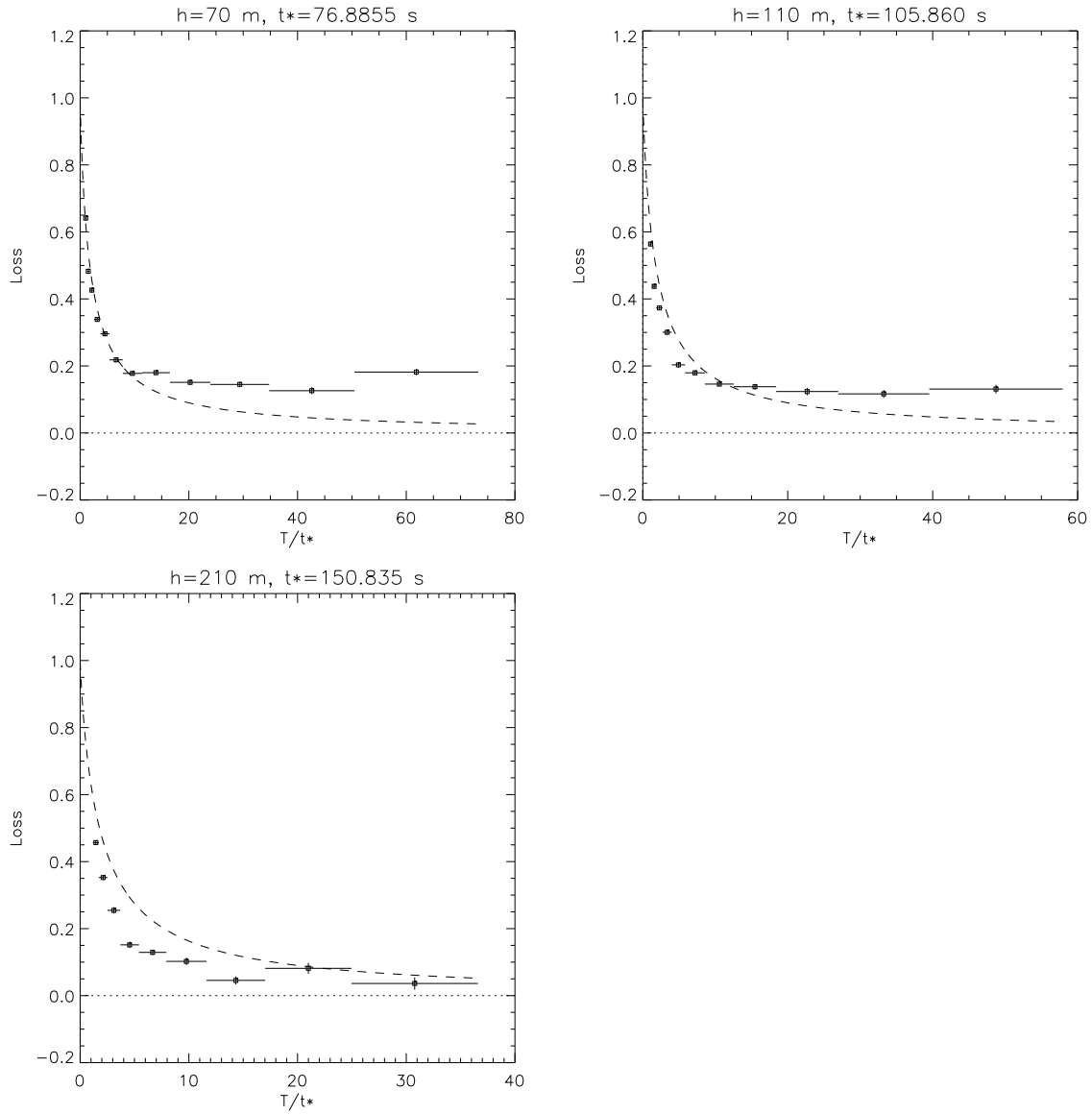
Figure 3.9: The behaviour of $\overline{w'T'}$ fluxes as a averaging time varies is compared with correction algorithms as derived from surface layer theory as figure 3.6, now based on LES results. 100 time series were randomly chosen from a single LES run from 4-5h simulated time. Black boxes represent the averages of EC fluxes as calculated from these time series, with respect to the corrected flux with $T_{max} = 10$min. The dashed line represents equation 3.19. Titles indicate the average values for the corrected t^*.

Figure 3.10: The behaviour of $\overline{w'T'}$ fluxes as a averaging time varies is compared with correction algorithms as derived from surface layer theory as figure 3.6, now based on LES results. 100 time series were randomly chosen from a single LES run from 4-5h simulated time. Black boxes represent the averages of EC fluxes as calculated from these time series, with respect to the spatial flux. The dashed line represents equation 3.19. Titles indicate the average values for the corrected t^*.
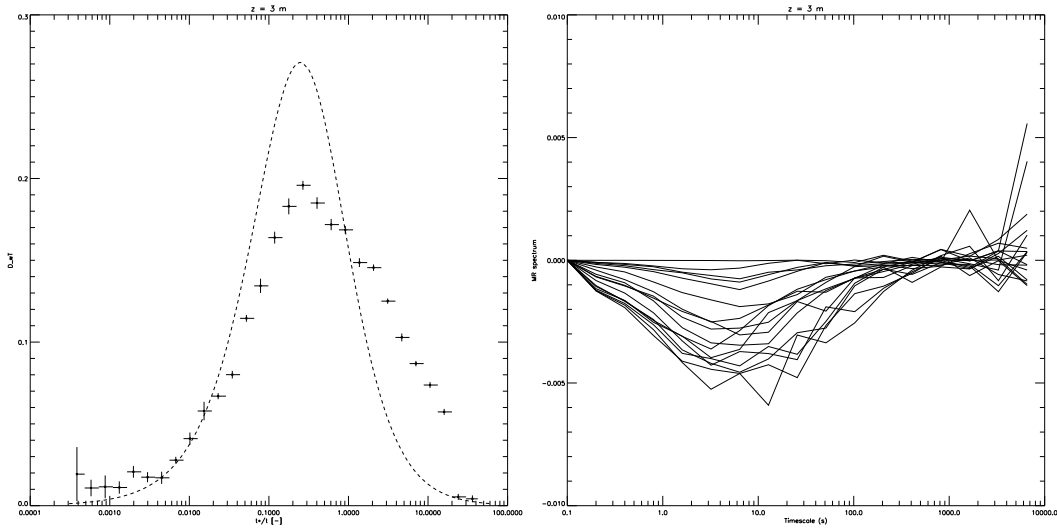
(a) Unstable data, scaled according to equation 3.26 and bin-averaged as before.

(b) Stable data, individual MR spectra without any scaling.

Figure 3.11: The behaviour of the MR spectrum based on 3m sonic data taken from three-hour periods from a) 10.30-13.30h and b) 0.30-3.30h.

With $t = t(m)$ as above. Now $F_C = \int_0^\infty S_{wc}(f)df \sim \int_0^\infty D_{wc}(m)/f df$, thus equation 3.25 becomes:

$$\frac{D_{wc}(m)}{\int_{-\infty}^\infty \ln(2) D_{wc}(m) dm} = \frac{4t^*/t}{\left(1 + \frac{4t^*/t}{p-1}\right)^p} \tag{3.26}$$

Figure 3.11(a) compares the 3m data to this equation as an example. Indeed, figure 3.11(a) closely resembles figure 3.1.

The increased amount of noise in figure 3.11(a) is due to the fact that a MR spectrum based on 3 hours of 10Hz data is composed of only 16 points, as timescales range from $t = \delta t = 0.1$ second to $t = \delta t 2^{16} \approx 1.82$ hours. Thus, no segmentation of the dataset is done, as was performed in the Fourier analysis. This is the reason larger timescales are reached than the Fourier analysis reached before. The disadvantage is that large jumps are made in timescale stepping, which introduces large errors when, for example, numerically integrating the spectrum.

However, it is not the idea of the MR spectrum to make it look like a Fourier spectrum. Instead, Vickers and Mahrt state that a gap region becomes clearly visible in the MR cospectra. This gap region is the region which is then supposed to seperate turbulent fluxes from mesoscale motions.

Indeed, figure 3.11(b), composed of seventeen three hour periods of stable periods after tilt correction, indicates the existence of such a gap time scale. Less days are used than the 32-day dataset for clearness of presentation, since the spectra are not suited for averaging. Scales of turbulent transport at night are generally much smaller than those in convective circumstances, however, so it is interesting to see whether a gap timescale can be seen for those circumstances.

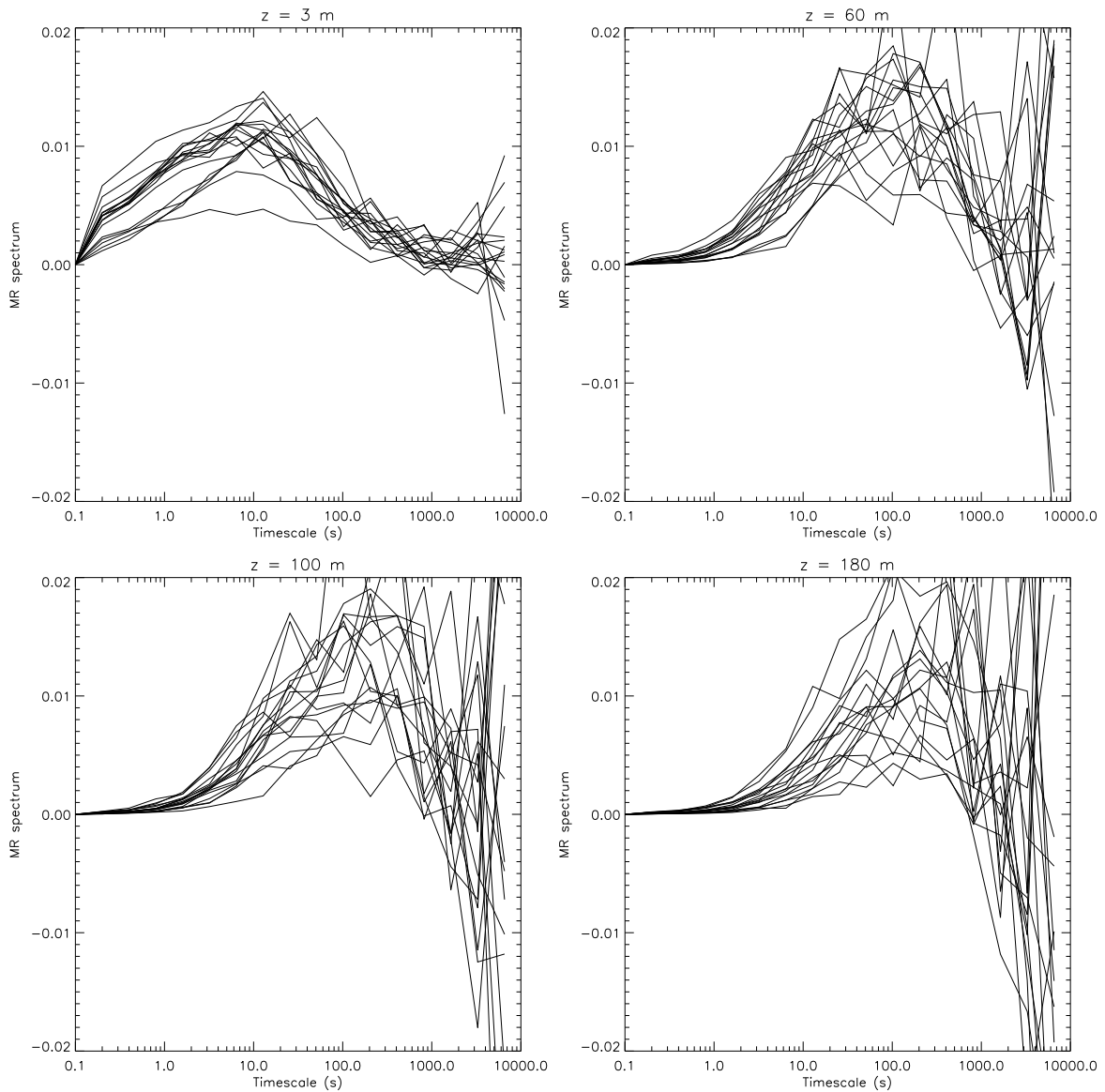Figure 3.12 shows the MR spectra of the dataset of seventeen days with three-hour periods

Figure 3.12: MR heat flux cospectra of seventeen three-hour periods around noon at several heights. No normalisation has been done.

around noon. While a gap timescale can mostly still be identified at 3m, the spectra become much noisier at greater heights and a gap timescale is not easily identified.

Vickers and Mahrt proposed the use of an automated algorithm to objectively find the gap timescale. This algorithm scans the MR spectra from small t to large after having applied a 1-2-1 filter to smooth out the noise. It then identifies the first peak. The gap scale is identified where the cospectra either increase or level off at a time scale greater than this peak time.

Note that the large amounts of noise at larger heigts will greatly inhibit the accuracy of this scanning algorithm, as it is clear that the noise is too large to be smoothed out with a 1-2-1 filter. This is identified as a random sampling error by Vickers and Mahrt, who propose to make use of an average gap scale over the entire dataset, instead of using the individual gap timescales as identified by the algorithm per time series.

It can also be observed that the amplitude of mesoscale contributions increases with height heights, and that both negative and positive mesoscale contributions occur.

Evaluating this averaged gap scales, Vickers and Mahrt seperate unstable, stable and near-neutral situations and find a clear height-dependence. In order to accomodate a comparison between the CASES99 data from Vickers and Mahrt and the Cabauw data used in this paper, the same distinction will be made here. Data with $|L| \geq 100$ is grouped as near-neutral, while data with $|L| < 100$ is grouped by sign into unstable and stable groups.

Data is selected from the 32 day data set as mentioned before. To allow for fair comparison with Vickers and Mahrt, one-hour periods are used. Each day two stable one-hour periods from 1.00-2.00h and 2.00-3.00h are selected, as well as 3 unstable periods from 10.30-13.30h. This gives a respective spread of 17, 67 and 76 one-hour periods for the near-neutral, unstable and stable groups. Figure 3.13 shows this data, where the data is averaged 'geometrically', hence the average of the logarithm of the values is taken, which corresponds to the logarithmic horizontal axis and gives less weight to large values, hence decreasing the spread.

Note that even with so much data and averaging logarithmically, the error in the averages is still very significant, and far larger than was observed in figure 3.3. This is due to low amounts of points per spectrum, which results in the fact that each single step upwards in gap timescale is a doubling of value. Furthermore, the strange behaviour of sudden decrease of gapscales for stable situations above 60m is caused by high amounts of noise in the MR spectra as the flux magnitudes approach zero for these situations. At this points the automated gap scale detection algorithm has severe difficulties pinning the right gap scale.

Quantitatively, the comparison is further made difficult by the fact that Vickers and Mahrt include heights from 0.5m up to 60m, which means that only the lowest two points of the Cabauw data, 3m and 60m, can be compared to their data. Note that this also qualitatively gives a distorted picture as they observed a fast increase of gap scales from 0.5m to 10m, with only three measurement heights in this range.

This increase is not witnessed on the Cabauw site, but this could be due to the fact that measurement starts at 3m and no measurements are done at 10m. At the comparable heights, Vickers and Mahrt observe typical gap timescales in the order of 200s for stable situations, while this grows to the order of 1000s in unstable circumstances. In this range, figure 3.13 displays timescales of the order of 100s for stable situations and 300s for unstable cases. This accounts to differences of a factor two to three. It might be that this is caused by differences in location as the gap timescale is expected to be dependent on mesoscale transport scales. This dependence on mesoscale motions might also explain the fact that figure 3.13 shows
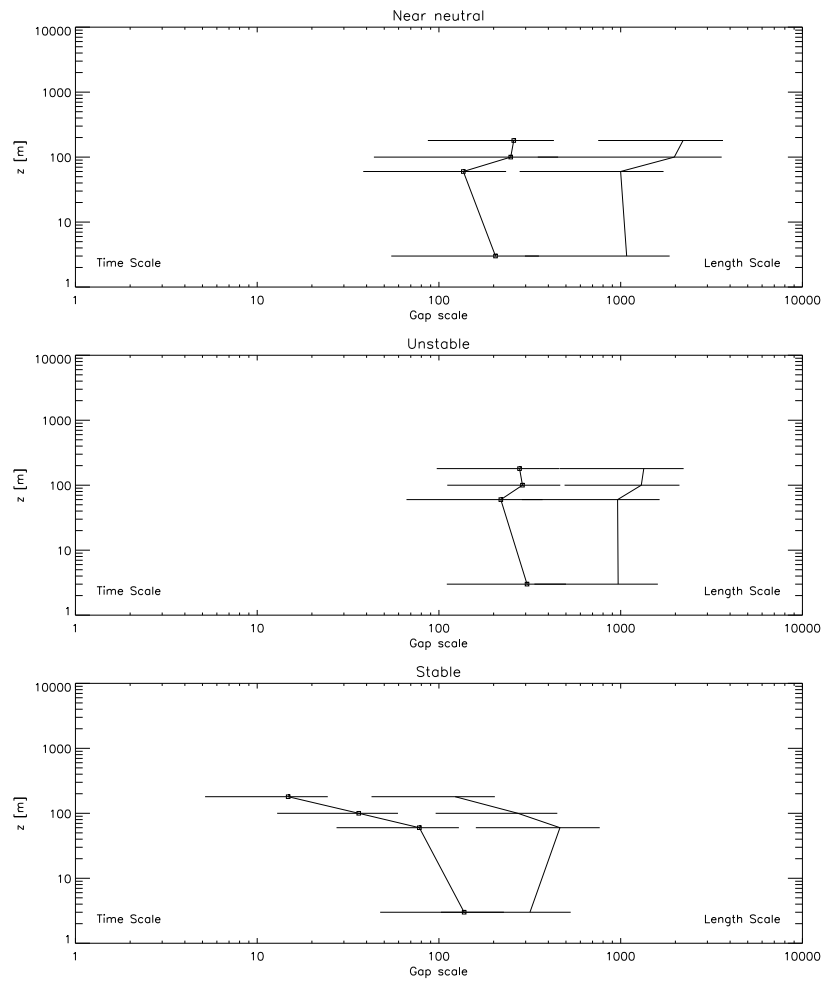
Figure 3.13: Height dependence of the gap scale (timescale in seconds, length scale in meters) as found by an automated algorithm as described by Vickers and Mahrt. The scales are found by logarithmic averaging.

very little height dependency, in contrast to figure 3.3 which is expected to show timescales of dominant turbulent transport. This clearly illustrates the difference in physical meaning of t^* and t_{gap}.

Drawing further on this interpretation, it is very interesting to note that features of figure 3.12 can be explained by comparing t^* and t_{gap}. It is clearly visible in the MR spectrum of figure 3.12 that while at 3m height a gap scale can easily visually be identified, this becomes increasingly difficult at greater heights. This can be directly related to the increase in t^* shown over this heights, where it reaches a value of the order of 80 seconds in unstable surroundings at 180m. Since the peak of the spectrum is reached at $4t^* \sim 320$ seconds, and it is seen that at this point $t_{gap} \sim 300$ seconds, this predicts the overlap between turbulent and mesoscale motions which can be directly observed in the MR spectrum of figure 3.12.

This illustrates well the problem of Eddy Correlation flux measurements, as time-based measurements of turbulent fluxes is obviously problematic when timescales of turbulent and mesoscales motions overlap.

Therefore, it seems this problem cannot be solved so simply by using similarity relationships. A localized view of the flux behaviour is necessary to further study the influence and behaviour of the different scales of motions in the atmosphere.

Chapter 4

Identification and study of thermal structures

As mentioned before, Kanda (2006) has suggested the presence of TOS to be (partly) the cause of the presence of flux on larger timescales and hence the resulting low frequency loss. Their LES studies showed coherent structures in all conditions. Figure 4.1 illustrates their results. It shows velocity maps of their LES studies for wind speeds from 0 to 4 m/s, averaged over 1 hour. Note that the average profile blurs somewhat as wind speed increases, but remains visible even after 1 hour averaging. This means results of one-hour averaging in timeseries will still be dependent on measurement location even with geostrophical wind forcing of 4 m/s.

In this section, an attempt will be made to identify and study this structures as they appear in time series. The same time series resulting from measurements at Cabauw will be used in the first instance, and will later be supplemented by LES results. The underlying idea is that with nonzero wind speed, these structures should propagate over the measurement point. Hence, the measurement in time can be seen as a one dimensional 'slice' of space. First, an algorithm is descibed, used to identify the thermal structures and their sizes as they pass.

The most basic unit of the larger thermal structures is the thermal, or plume. In Stull's *Boundary Layer Meteorology* (1988), thermals are defined as 'large columns of rising buoyant air in the convective mixed layer'. They are differentiated from plumes, which are defined as 'vertical structures of warm rising air', which does not seem to be a definition which is very different from the thermal. Therefore, since the main process in both structures is the upward transport of warm air, and since both structures have sizes of the same order (hundreds of meters to the order of a km), in this report no difference will be made between the two.

In order to identify these thermals, commonly attempts are made to use criteria which label specific points of the time series as a thermal. De Haij (2005) also experimented with the use of wavelet analysis to visualize the structures.

In this report, the criteria will be chosen such as to require the presence of warm, rising air with an additional requirement of vertical coherence, as deduced from the definition in Stull. One of the unique features of the Cabauw measurement tower is the presence of high-frequency sonic anenometers at a range of different heights. Using the upper three heights at 60m, 100m and 180m and excluding the lowest 3m measurement height as at this height
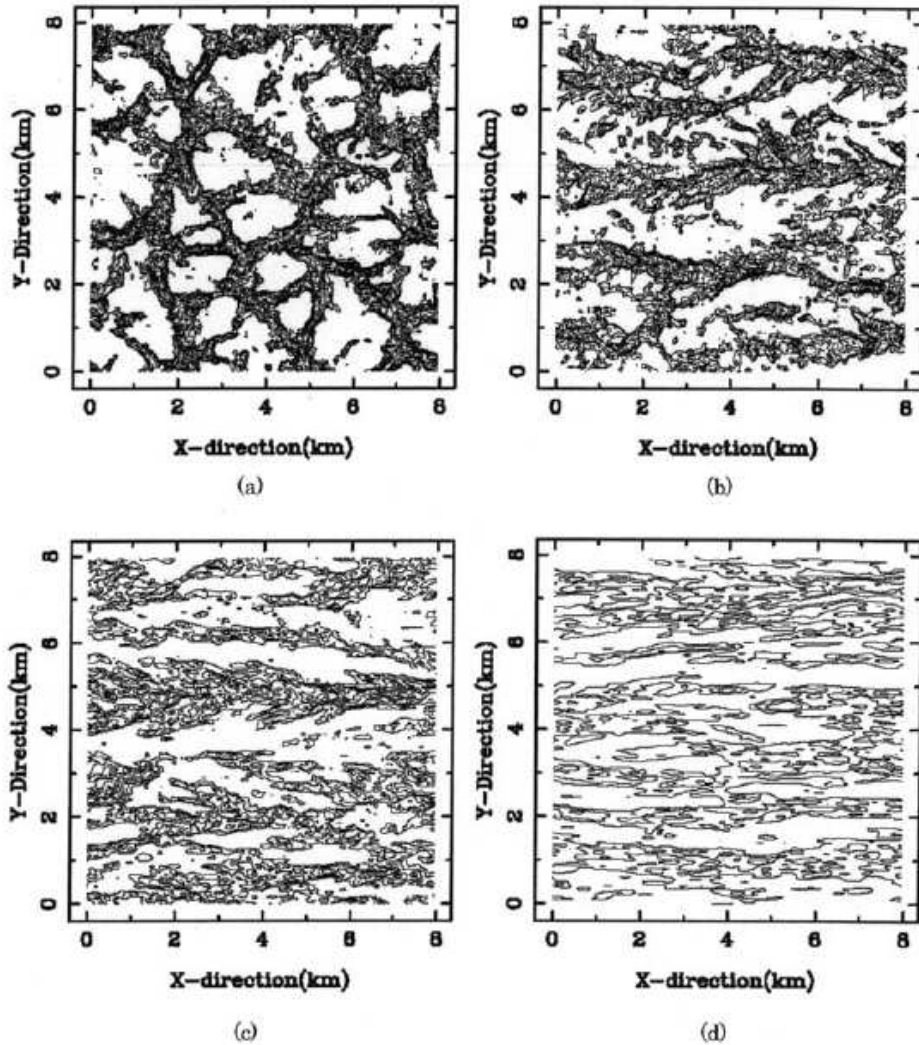
Figure 4.1: Figure from Kanda et al (2006). Vertical velocity maps at 100m height, averaged from 2 to 3 hours simulated time. The figure shows velocity maps for geostropic winds of a) 0 ms^{-1}, b) 1 ms^{-1}, c) 2 ms^{-1} and d) 4 ms^{-1}
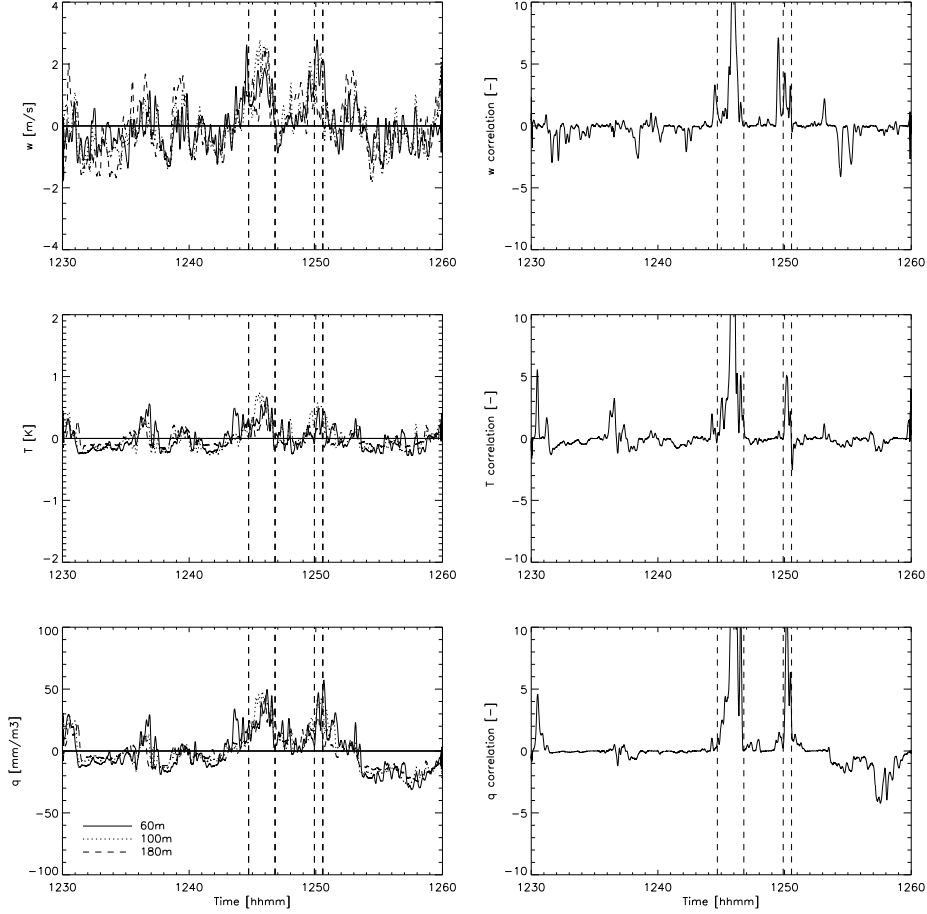
Figure 4.2: Figure illustrating the use of the given criterea on the correlation 4.1 Left the time series of w', q' and T_v for the different heights can be seen, while on the right the correlations are visualised. The dashed lines represent the thermals as marked by the criterea. The time series is passed through the same triangular filter as the correlation in order for visualisation purposes. Data is used from 12.30 to 13.00hr of May 10th, 2008.

no detectable structure is expected, the vertical coherence is tested using the three-height-correlation of vertical velocity as follows:

$$\rho_w(t) = \frac{w'_{60m}(t)w'_{100m}(t)w'_{180m}(t)}{\sigma_{w60}\sigma_{w100}\sigma_{w180}} \tag{4.1}$$

The same correlation is calculated for the temperature T. These correlations are only expected to be significantly positive if a positive deviation from the mean is found at all three heights.

A triangular smoothing filter with a width of 1.5 minute is used on the resulting correlation series, which effectively removes all uncorrelated variations. A thermal is now identified as those points in time in which $\rho_w(t) > 0$ and $\rho_{Tv}(t) > 0$ *for more than 95% of the time in an interval with a minimum width of τ.*

The resulting process is illustrated in figure 4.2, where the time series at all heights can be seen on the left and the resulting filtered correlation on the right. Dashed vertical lines

illustrate begin- and endpoints of the marked thermals. Note that the differences between turbulent motions and coherent vertical structures present at all three heights is clearly visible in the correlated time series. For the figure a time series of May 10th, 2008, of 90 minutes starting from noon, is used. The detection is done with $\tau = 50s$ for this figure so as to clearify the process. Further in this report, the detection window is always $\tau = 15s$ (following de Haij, 2005).

However, while the presence of thermals can now well be identified, figure 4.2 also clearly illustrates the diffulty in defining the edges of the thermals. Looking at the time series, it would be very hard for a human being to pinpoint the edges - and hence the size - of the thermal structures, and the given criteria do no better. The size as found by the algorithm is somewhat dependent on the minimum-time criterea of τ, which is clearly an undesirable situation.

This difficulty in defining the size is not a surprise, as Stull already visualised a cross section of a thermal as having a circular, but cloud-shaped form as shown in figure 4.3, with an inner core and an intromission zone around. Within the intromission zone, the thermal is less well-defined and hence a lot of noise will be introduced at the edges.
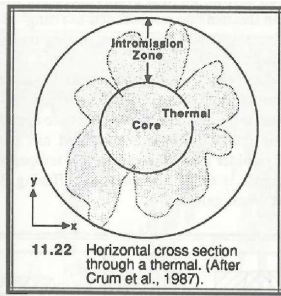

Figure 4.3: A visualisation of a horizontal cross section of a thermal from Stull's *Boundary Layer Meteorology* . The figure clearly illustrates the difficulty in defining the thermal's edges.

However, a study on time and length scales in the atmosphere still requires an objective classification of the size of a thermal. This is achieved by introducing a second algorithm. This algorithm is an extension of the first one. It makes use of the wavelet transform (see section 2.2) to find under which scale the greatest power is found. The algorithm considers a part of the wavelet transform of the vertical velocity w', 3 minutes wide, centered around the thermal as detected by the first algorithm. It creates a wavelet power spectrum of this *local part* of the wavelet transform, which it smoothens with a 1-2-1 filter. Now the scale $s = s_{max}$ is calculated for which the wavelet power is at its maximum.

The algorithm then considers the wavelet transform at this scale, $W(t, s = s_{max})$, which is in essence the time series filtered with the wavelet mother function with scale s_{max}. This series is now used to find the exact location of the centre of the thermal by determining the location t_{max} of the local maximum in this time series. The detected thermal is now defined as the area $t_{max} - s_{max} < t < t_{max} + s_{max}$ (remember that the Mexican Hat wavelet function has zero crossings at $-s$ and s).

The result of the second algorithm is illustrated in figure 4.4, where the time series is displayed together with it's wavelet transform and the global power spectrum. Dashed vertical lines indicate the beginning and end of the detected thermals as before. Note that the larger
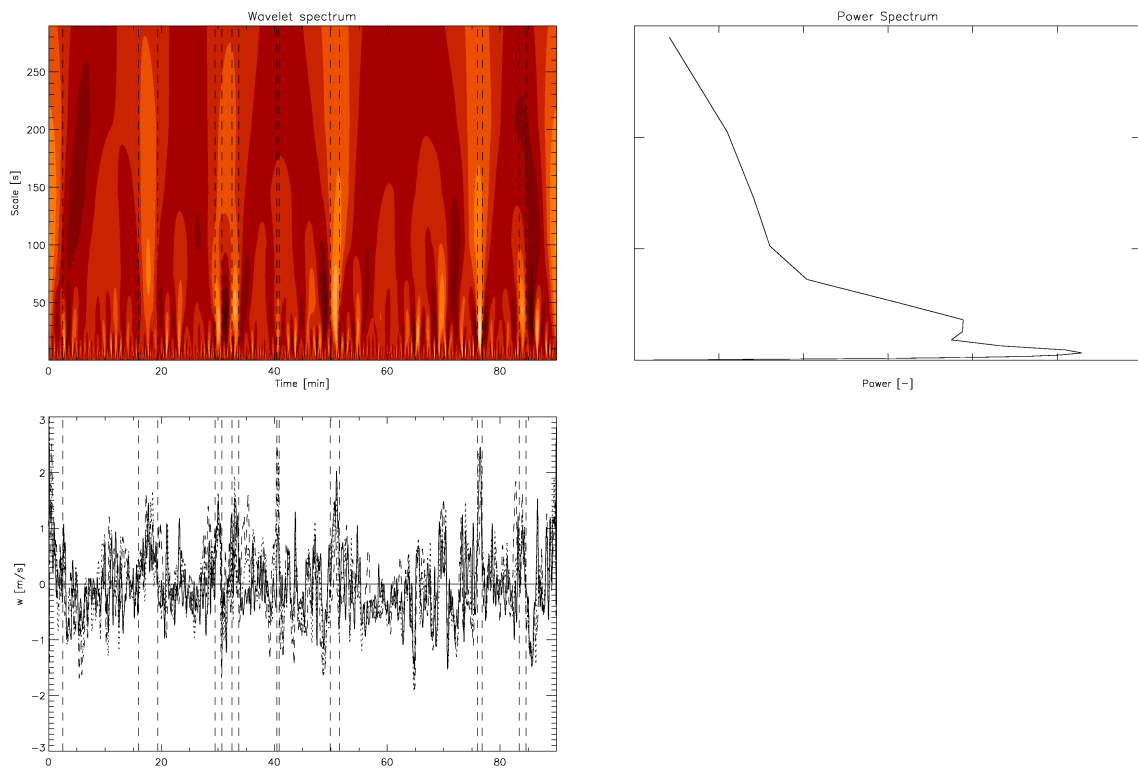
Figure 4.4: Figure illustrating the work of the algorithm used to determine the edges of the thermals. Vertical dashed lines denote the edges of detected thermals as before. The wavelet is shown of a 1.5hr interval starting at noon of May 6th, 2008.

structures are beautifully visualized in a wavelet spectrum, being the remaining peaks as the wavelet scale increases. However, the scale of the peak in the global power spectrum is usually much smaller than the scales of the thermals, indicating small-scale turbulent motions dominate this global spectrum. This is the reason why the local power spectrum is regarded in the classification of thermals.

The typical characteristics of the points which are detected as thermals are visualized in probability density functions of vertical velocity, moisture and temperature at the three upper heights in figure 4.5. The shown pdfs are based on the full dataset (24 time series; the data for which the inversion layer height could be profiled) equal to what will be used for later figures.

The solid line represents the total probability density function of the data, while the red line indicates the part of this function which is identified as being a thermal. The dotted line is the complementary probability density function of the non thermal part. The thermals clearly occupy the larger positive parts, but perhaps not so clear as would be expected. However, the pdfs are expected to be smeared out due to the spread in behaviour of the different days. The thermal probability density functions do seem to have a gaussian shape themselves, especially in vertical velocity. This is a comforting result as this is often assumed in modelling.

The use of the wavelet algorithm to improve the exact size of the thermals has two large advantages. First, it provides a more objective, parameter-indepent way of defining the exact scale, while not interfering with the requirement that the thermal has to stretch over all heights by the use of the first algorithm. Second, it provides a way of pinpointing the exact location of size of the thermal at different heights when executed over the w' time series at the local height. This allows a study of the shape and angle of the thermals as they pass over the measurement site.

Figure 4 shows contour plots of vertical velocity w' as function of height and time. On the contour plot, the edges of the detected thermals are plotted in white, illustrating the vertical structure. Only thermal structures lasting longer than 60 seconds are visualized for clearity. This is the reason only a part of some thermals are visualized, as the not-shown part then shrinks to a width below 60 seconds.

The contour plot is not very detailed as it based on only four timeseries between 3 and 180m, giving an exaggerated sense of vertical coherence as the contour algorithm interpolates data in between the measurement levels. However, the thermal structures are still visible as relatively broad, light (high w') vertical bands, and are well detected by the algorithm. The algorithm is not used to detect thermals at 3m, since at this height they are so diffuse they cannot be distinguished from turbulence.

Note that as the horizontal scale spans 90 minutes while the vertical scale spans only 200 meters, the contours have an aspect ratio of about 1:135, assuming an average wind speed of $5m/s$. Interesting is that while too diffuse at 3m, the thermals are already well defined at 60m, and change little in width as the height increases up to 180m. It would be interesting for a possible future study to see from which height the thermals can be detected and see how the structure changes when nearing the ground.

The figure also shows how the thermals pass the tower at all three heights. This illustrates the concept of large vertical columns of air which move as a coherent structure through the atmosphere, in contrast to buoyant parcels of air which move upwards in a bubble like way. Note that The wavelet analysis enables us to reconstruct the angle of the thermal with re-
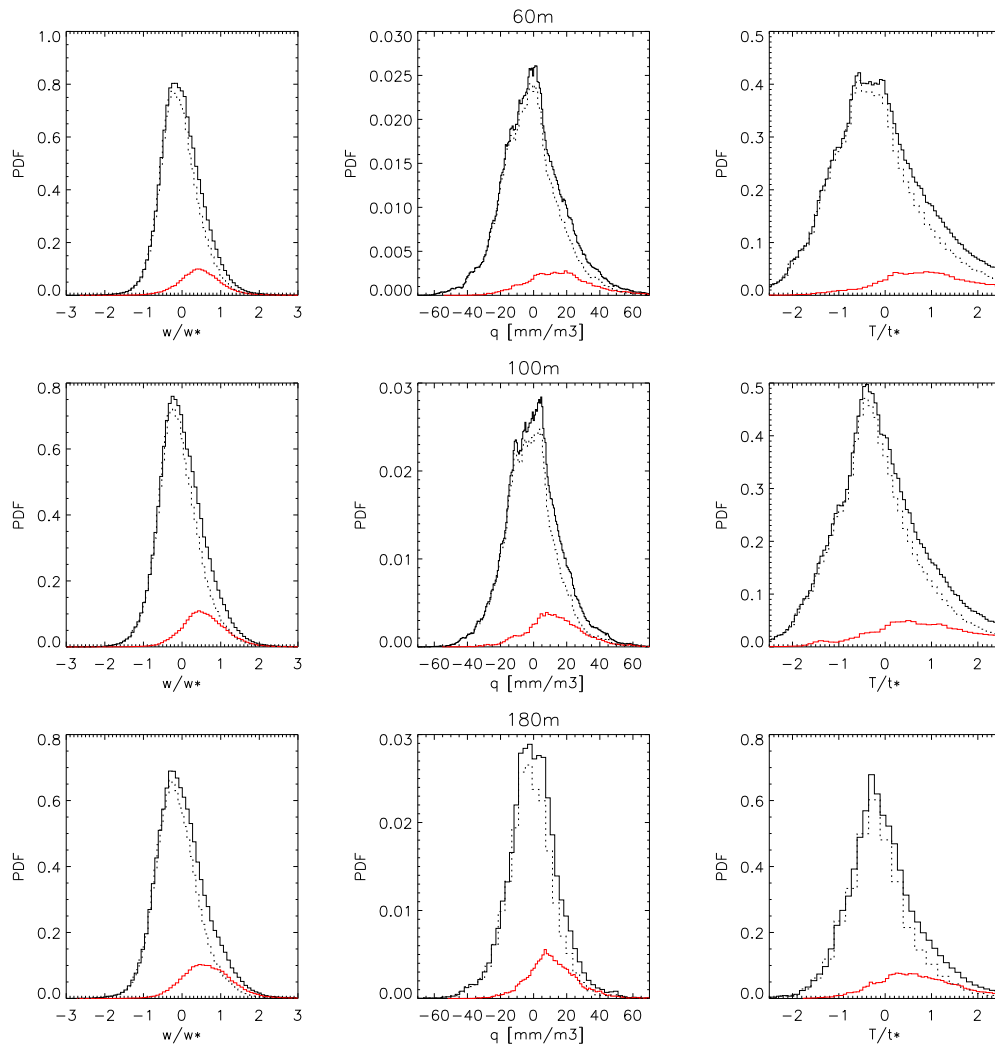
Figure 4.5: Probability density function of the characteristic properties of the the thermal structures based on 24 time series of 1.5 hour starting at noon. Solid black lines represent the complete PDF, while the red and dotted lines represent the thermal and non-thermal parts of the PDF, respectively.
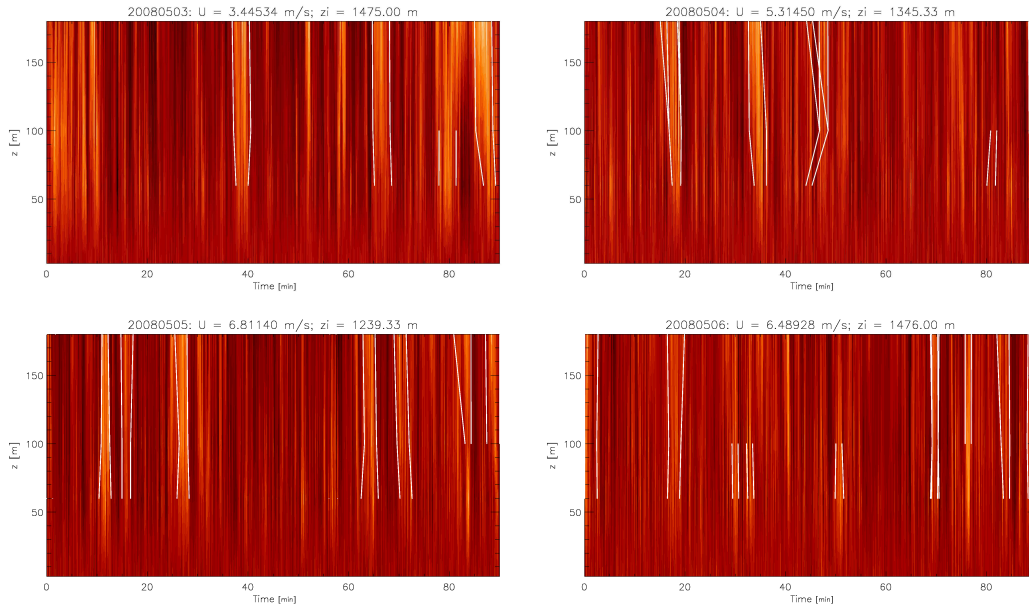
Figure 4.6: Contour plots of vertical velocity w' as function of height and time for the period from 12.00 to 13.30h, based on sonic data from 3, 60, 100 and 180m. White lines are reconstructions of the thermal shape based on the size and location of the detected thermals at the upper three heights.

spect to the ground. Stull describes that wind shear causes the thermals to move in a 45 degree angle in the direction of the propagating wind speed. This would result in the thermal arriving 180m earlier than 60m, and hence a time delay would be detectable.

Thermals propagate with a speed equal to the averaged wind speed over the vertical (Stull). Therefore, the used wind speed is an average over data from the four used heights. Using this wind speed and the detected time delay, the angle can be calculated based on the difference between the centrum of the thermal at 180m and 60m, using Taylors hypothesis of frozen turbulence to transform the time delay to a length scale.

Figure 4.7 shows that this is hardly a constant factor - a wide range of angles has been found, with no clear dependence on wind speed itself. In fact, it is even possibe for thermals to be tilted backwards (against the wind). It is expected that the thermals are much more complex in shape than simple cilinders, merging with each other and changing shape as height or time changes. This complexity is the expected cause for the wide range in angles. However, there is a clear peaking of the probability density function in the range of very large angles. This is currently unexplained and it remains a question whether this is a physical phenomenon or an artifact of the attempt to reconstruct the thermal.

Figure 4 shows how the mean size of the thermals increases with height. Data is grouped by inverion layer height. As expected, a considerable increase in size is found. In fact, the mean size nearly doubles as height increases from 60 to 180 meters for high inversion heights. Interesting is that while a significant difference of size is found between the lowest and highest inversion classes, no mentionable difference is found between the higher classes. This might indicate the lowest class is in a different regime.
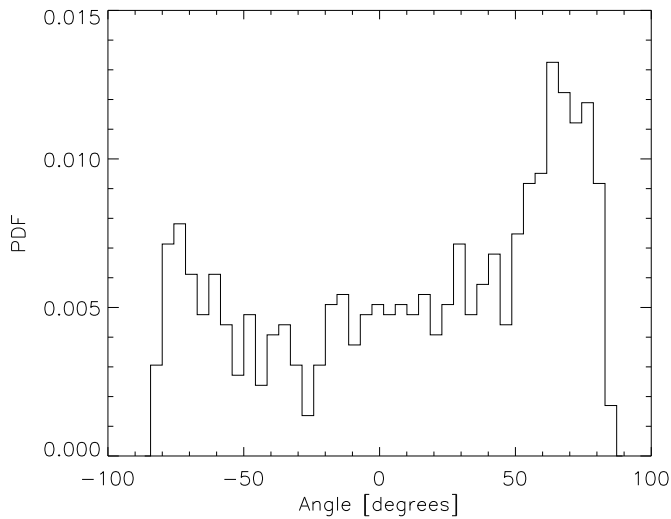
Figure 4.7: Probability density function of the angle the vertical thermal structures have with respect to the vertical, with positive angles indicating a tilt in the direction of the propagating wind speed.

Returning to the time series, figure 4.9 shows the mean shape of the vertical velocity anomality and temperature anomaly induced by the detected thermal structures as they appear in the measured time series. Only time series dataset from section 2.1.1 measured in spring and summer are used such to have a dataset as homogeneous as possible while still having enough data to average. In this dataset, 2853 thermals were detected. The thermals show a wide diversity in shape, but averaging these thermals over an horizontal axis which is normalized with the detected size reveals the common shape. The part of the given shape which is classified as thermal is thus plotted from -0.5 to 0.5. This also illustrates and explains the earlier given probability density functions, which show a tail of both w' and T' towards zero, which seems to be caused, at least partly, by the edges of the thermals.

As the horizontal axis is a rescaled time series, the leftmost points are points detected first, and hence the thermal moves with governing wind, directed from right to left.

Vertical bars denote the standarddeviation σ (not to be confused with the error *in the mean* of equation 3.12). The general shape of the pattern is remarkably similar to the Mexican Hat (section 2.2), indicating the right choice was made with respect to the wavelet's 'mother function'.

A very remarkable feature of the figure is the 'dips' before and after the thermal, which show that right before and right after the thermal, descending cold air is found. A slight asymmetry is witnessed in this shape as well, as the 'dip' *after* the thermal (hence trailing the thermal) seems deeper and more steep. This dip is typically around $1/3^{rd}$ as deep as the peak is high for vertical velocity, while the temperature dip is less pronounced. Interesting is also the fact that the structures start and end somewhat above the zero line, indicating an overshoot when returning to zero. This is unlike the Mexican Hat shape, and it means that they will have to go below the line further on, in order to sustain the average. This is in contradiction with the concept of large thermals above the zero mean with large negative areas around them.

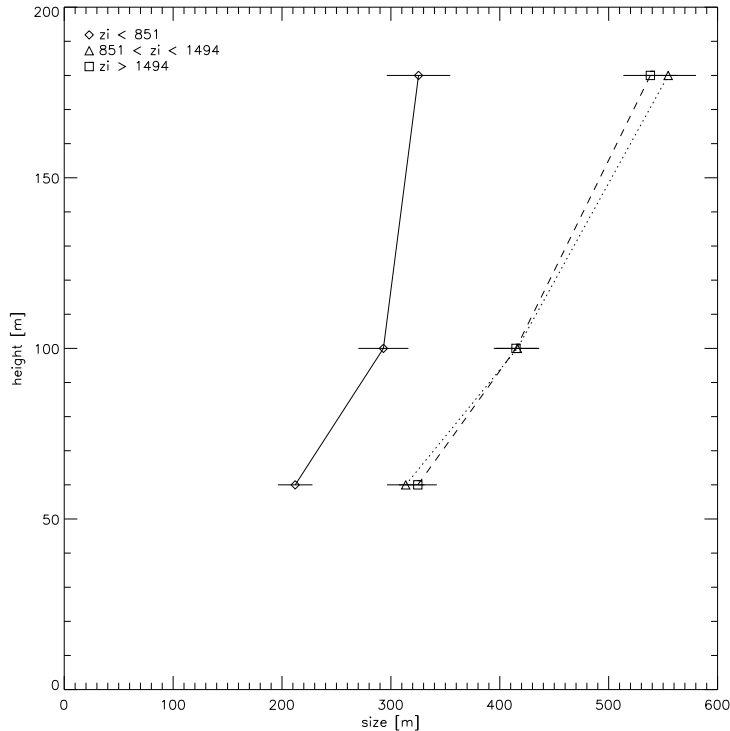It is also interesting to see how the maximum velocity anomaly increases with height, while

Figure 4.8: The behaviour of the mean size of the thermals as height increases. Horizontal bars indicate the standarddeviation in the mean. Only data from May 2008 is used to reduce surrounding influences. Data is grouped by ranges of inversion height z_i.

the temperature anomaly decreases with height. This is in accordance with expectations, as measurements are done in the lower part of the boundary layer where the air is still accelerating upward. As the mean temperature increases with height, the temperature excess of the thermal decreases with height.

Now that some of the qualitative features of thermal structures have been discussed and visualized, it is interesting to see whether data from the thermals as detected by the algorithms can be used to make quantitative statements on the behaviour of thermal structures.

Proceeding along the lines of this report, it makes sense to start with the timescales involved in thermal transport, or more accurately, the width of the detected thermals. This is expected to be hard, as the position of the measurement tower in itself already produces a spread in the size of thermals. Luckily, as is discussed in appendix A, the probability density function which is introduced by this phenomenon increases relatively sharply towards the true thermal size. With enough statistics, the behaviour can still be analyzed.

Towards this end, it is useful to introduce Chebyshev's inequality for later use. This inequality allows one to make quantitative predictions on the possibilities of an event happening, even if no information is known a priori. Chebyshev's inequality is as follows (Dekking et al, 2005) *For an arbitrary random variable Y and any a > 0:*
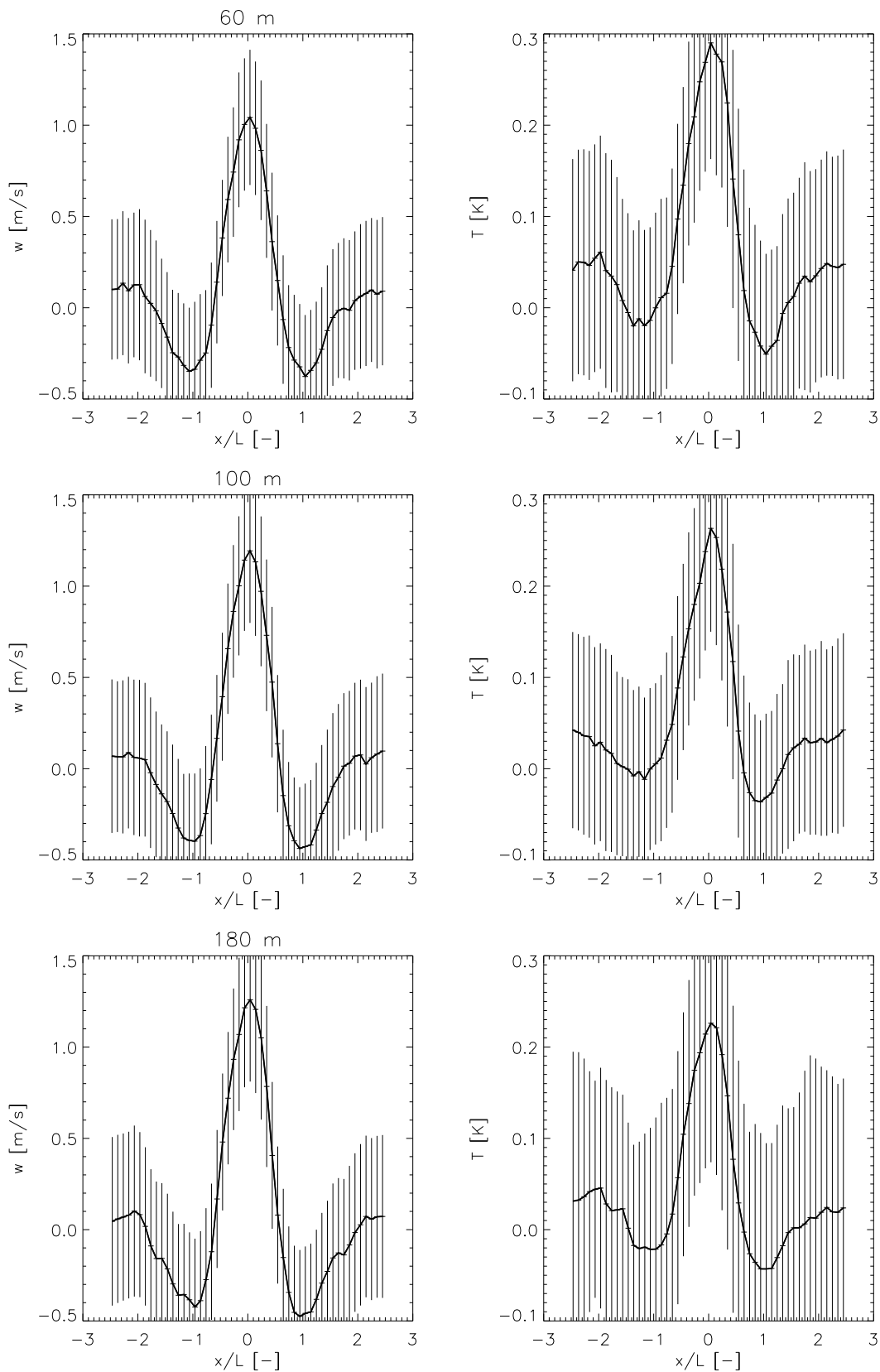
Figure 4.9: The normalized shape of the time series anomalities induced by thermal structures as they pass over the measurement site. The horizontal axis is normalized over the detected size, hence the part of the shown structures which are classified as thermal ranges from -0.5 to 0.5. Vertical bars denote standarddeviations.
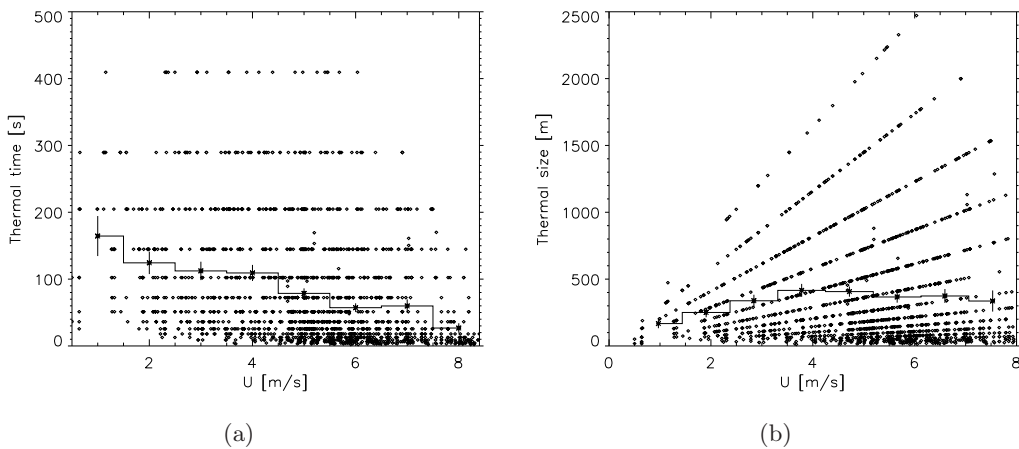
(a)   (b)

Figure 4.10: a) The time a thermal takes to pass over the measurement site T and b) the calculated size, UT, are plotted as a function of the 10-minute propating wind speed U. The solid line represents the averaged scale with vertical bars indicating the conservative 75% confidence interval.

$$P(|Y - E[Y]| \geq a) \leq \frac{Var(Y)}{a^2} \qquad (4.2)$$

Hence: the probability of an error in the evaluation of the mean being larger than a, is smaller than the variance over a^2. This is a strong basis for quantitative predictions as the variance of the mean decreases with $1/N$.

Turning back to the data, those days of the 32 day dataset are used for which the boundary layer height z_i could be profiled. For these days, four 1.5 hour periods per day from 9.00 to 15.00h were analysed and data from detected thermals was saved. This yielded 1008 detected thermal structures.

Figure 4.10(a) shows the timescale data as a function of the propaging wind speed U. The use of the wavelet-based algorithm implies that the detected timescale follows discrete steps, following each other with power $\sqrt{2}$ (Single data points outside this range are thermals detected on the edge of a considered period). As no direct structure can be seen in the data, a first hypothesis would be that the timescale T is independent of propagation speed, indicating that thermal structures stretch out as wind speed increases.

To test this hypothesis, the solid line represents the averaged values over horizontal wind speed bins. Using Chebyshev's inequality with variance as the square power of the standard deviation of the mean, equation 3.12, vertical lines can be constructed as the 75% confidence intervals. It now becomes apparent that there is in fact, a drop in average thermal timescale is found as wind speed increases.

Indeed, averaging over only two bins, it can be found that if indeed the timescale was independent of wind speed, the chances of obtaining averaged T as far from the global mean as data from figure 4.10(a) is less than 1%. Hence it can be concluded that *as propagation wind speed increases, thermal timescales decreases* .

In fact, figure 4.10(b) seems to indicate the typical structure size (again calculated with the help of Taylor's hypothesis) itself is independent of wind speed. The figure shows the same timescale data, now multiplied with their propagation speed to find their spatial width. The averaged data seems to remain constant, as the initial increase for small U might be explained
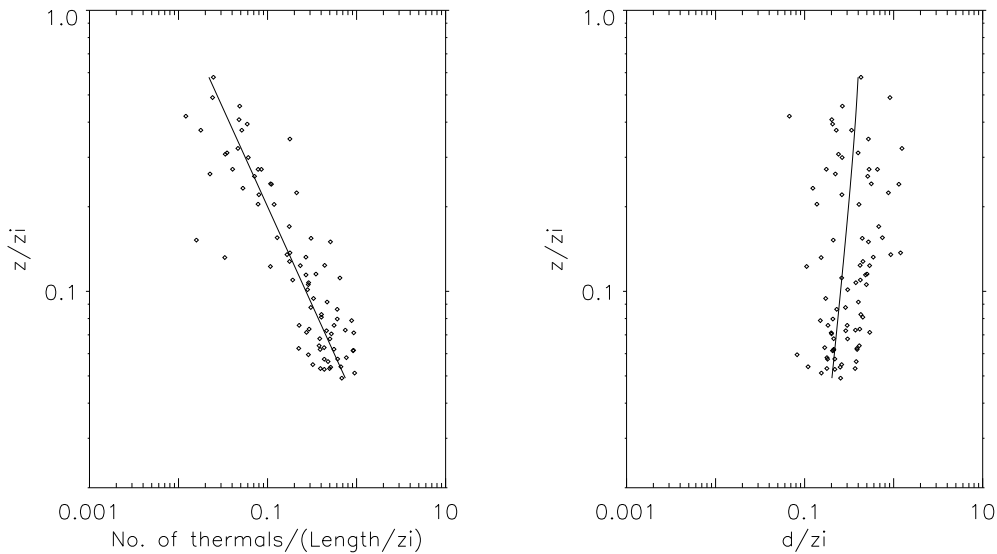
Figure 4.11: Number of thermals intersected per normalized length of timeseries and normalized average thermal length as a function of z/z_i, following Lenschow (1980). The solid lines in the left and right figure represent equations 4.5 and **??**, respectively.

by the fact that the top sizes are limited by the scale coverage of the wavelet transform. Indeed, no statistical evidence is found to reject a claim that *the size of thermal structures is independent of wind speed* .

Lenschow et al (1980), found that normalized thermal size to be solely dependent on z/z_i, where they have used data from overseas aircraft measurements at 135m to obtain:

$$\frac{d}{z_i} = 0.16 \left(\frac{z}{z_i}\right)^{1/3} \left(1 - 0.25\frac{z}{z_i}\right). \tag{4.3}$$

Here d is the average thermal length per measurement. Cabauw measurements as used in this paper had an average detected thermal size which is typically *3.5 times as large* , as can be seen from the solid line in figure 4.11(b), representing:

$$\frac{d}{z_i} = 0.56 \left(\frac{z}{z_i}\right)^{1/3} \left(1 - 0.25\frac{z}{z_i}\right). \tag{4.4}$$

The increased thermal size, as well as the increased spread in figure 4.11(b) relative to Lenschow's data can be attributed to a number of causes. First, Lenschow uses overseas aircraft data, while in this report thermals are measured using a stationary tower. Secondly, Lenschow uses a direct criterium (on humidity) whereas figure 4.11 has been created with the help of wavelet analysis. Thirdly, the days selected in this paper are completely cloudless, where Lenschow measured underneath shattered or broken stratocumulus.

This wide range of causes makes it impossible to point at the direct cause of differences. In fact, it is surprising that - apart from the factor 3.5 - the average size behaves so similar as a function of z/z_i. Indeed, figure 4.11(a) shows the normalized number of detected thermals, being the number of thermals divided by the measurement length over the boundary layer
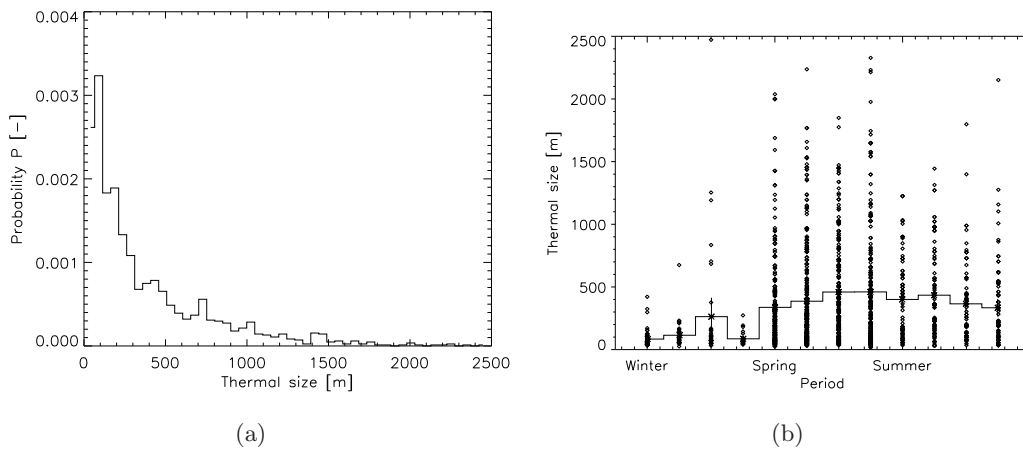
Figure 4.12: The distribution of thermal sizes. a) The probability density function of thermal size. b) The distribution per season and time of day, as data is grouped to season and time period, the four vertical columns per season being the four 1.5 hour segments per day, from morning(left) to afternoon(right).

height. This figure shows a much different behaviour as found by Lenschow: while Lenschow finds $N = 0.68(z/z_i)^{-1/3}$, figure 4.11(a) suggests a -1.43 power law:

$$N = 1.0 \cdot 10^{-2} \left(\frac{z}{z_i}\right)^{-1.43} \tag{4.5}$$

This difference may be attributed to the fact that in this report vertical coherence is demanded, which is expected to reduce the number of detected thermals.

Figure 4.12(a) shows a probability density function of the size of the intersected segments of thermals, based on all 1008 detected thermals. It shows very clearly that the distribution in size is not explained by the fact that segments will be intersected in various ways, since the pdf looks very different from the pdf calculated in appendix A.

Figure 4.12(b) shows how the distribution of size is influenced by the season. A clear increase in average size is found moving from winter to spring, but none from spring to summer. The distribution is relatively equal for morning or afternoon, especially in spring and summer months. This is interesting, as in this time a large increase in boundary layer height is witnessed. This indicates that boundary layer height is not the only relevant parameter in the process, at least at heights below 200m. It seems the full extend of the complexity of the thermal structures in the convective boundary layer cannot be caught in simple similarity relations.

4.1   LES study on thermal structures

For a better understanding of the thermal structures which have been identified in the measured time series, use is made of the LES model. The advantage of this is that the one

dimensional view of the time measurement series is now expanded to a full four (spatial and time) dimensional view. This allows us to study how time series relate to spatial behaviour. A dry LES run is done with 128x128 grid boxes on each vertical level. First, an equal analysis as done on Cabauw data is applied to LES results. 100 randomly chosen time series are extracted from the grid at 70m, 110m and 210m height, and in these series thermals are detected. Figure 4.13 confirms that the detected thermals have indeed the same shapes in the time series as was found in the Cabauw data. It shows that the acceleration and cooling of the air as it moves up, which was seen in the Cabauw data, continues as the height increases. The air starts to decelarate above the inversion, which is located between 200 and 300m for the given period, where the temperature anomaly becomes negative, exactly as expected.
The descending air on the edges of the thermals are also found in LES data, although the asymmetry is less obvious. It seems the temperature dip preceding the thermal evolves into the negative temperature excess above the inversion.

   Figure 4.1 gives an overview of the behaviour of the thermals as the air inside propagates towards greater height. The figure distinguishes cases with different inversion heights, indicating how the process is dependent on this height. It is clearly visible how the temperature excess within the thermals drives the air, accelerating up to the inversion, where the temperature excess dissapears and even turns into a deficiency, thus decelerating the air. This confirms theoretic expectations.
Figure 4.1 visualizes how the thermals grow with height. The functional form of the size is very similar to the vertical windspeed, growing up to the inversion after which the thermal starts shrinking.
Equation 4.3 predicted this qualitative behaviour quite well, but the quantitative behaviour is less well predicted, however. From lowest to highest inversion height, the dashed, dotted and dash-dotted lines indicate the behaviour of this equation after being multiplied with a factor of 14.9, 8.4 and 6.4, respectively. It seems a part of the behaviour is not yet caught within the dynamics of equation 4.3.
Comforting is that the LES results at lower heights behave much alike the observed time-series, where the sizes grow to the order of 400 meters as height increases up to 200 meters. This is comparable to the orders found in figure 4.

To study how the detected thermals in the time series relate to the thermal structures in space, the thermal detection on the basis of timeseries is extended to each of the 128x128 time series in the LES grid. These results can then be combined to give a 2D overview of where the thermals are located as a function of time at several heights.
Figure 4.17 shows the result of this method at a single time instant. Note how the detection algorithms, while not faultless, clearly detect the hexagonal structures which are expected. A very interesting result of this fact is that, as the algorithm requires, the same structures would be visible at roughly the same locations, in the vertical velocity and temperature field at greater heights (up to 200m).
   As time increases, the structures remain intact and pass over the grid. This allows the time series to pick up the thermals as they pass by, such that adjacent time series 'pass' the thermals over to each other. Imagining the structures of figure 4.17 moving in horizontal direction is thus very insightful when considering the pdf of figure 4.12(a). The thinner 'walls' of the structures pass over the measurement site in a multiple of angles, hence introducing a wide range of measured sizes. Besides that, it is clear that the thermals are not simple round
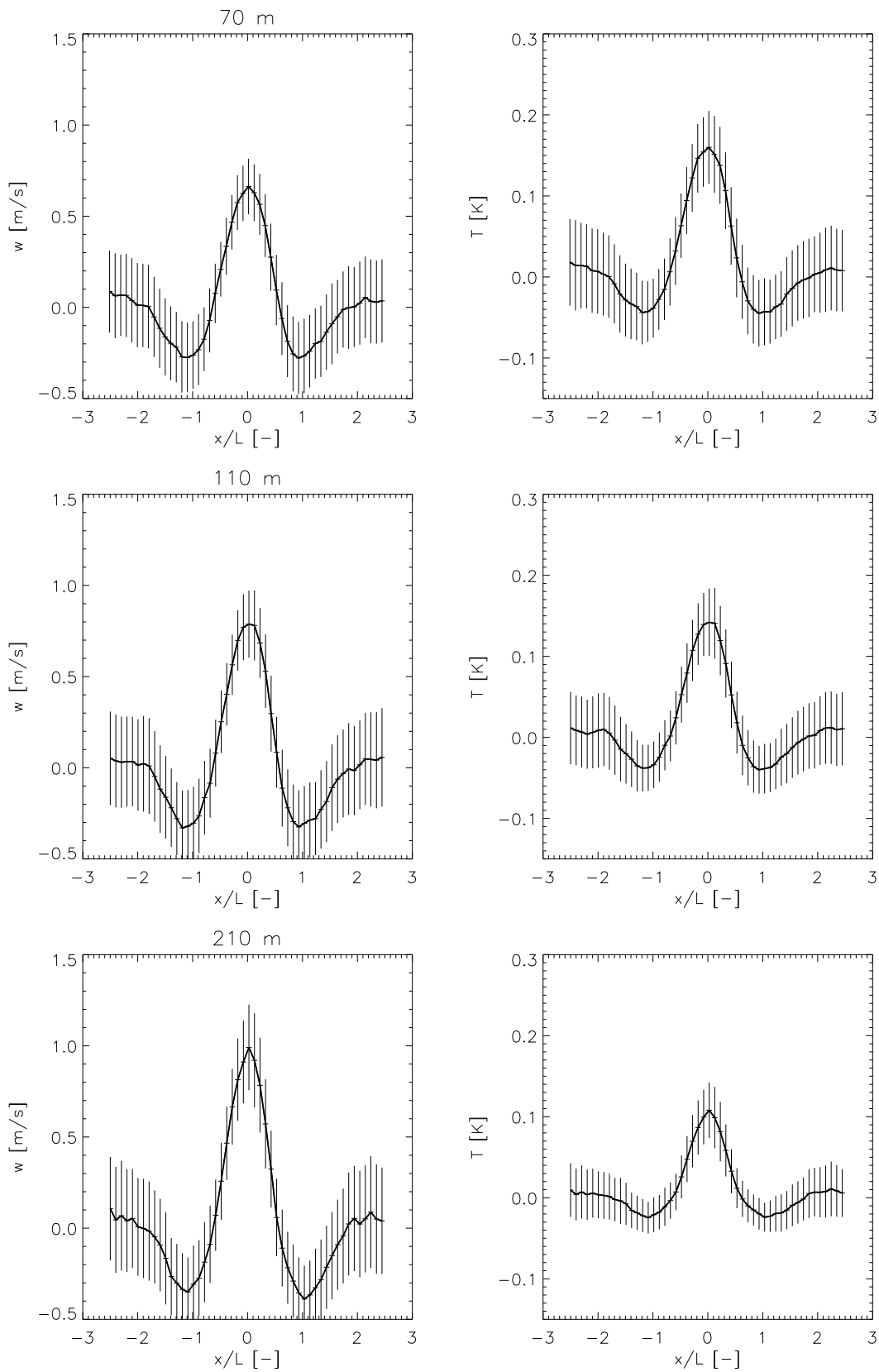
Figure 4.13: The normalized shape of the time series deviations induced by thermal structures as they pass over a virtual measurement site on a LES grid cell. The horizontal axis is normalized over the detected size, hence the thermal ranges from -0.5 to 0.5. Vertical bars denote standarddeviations.
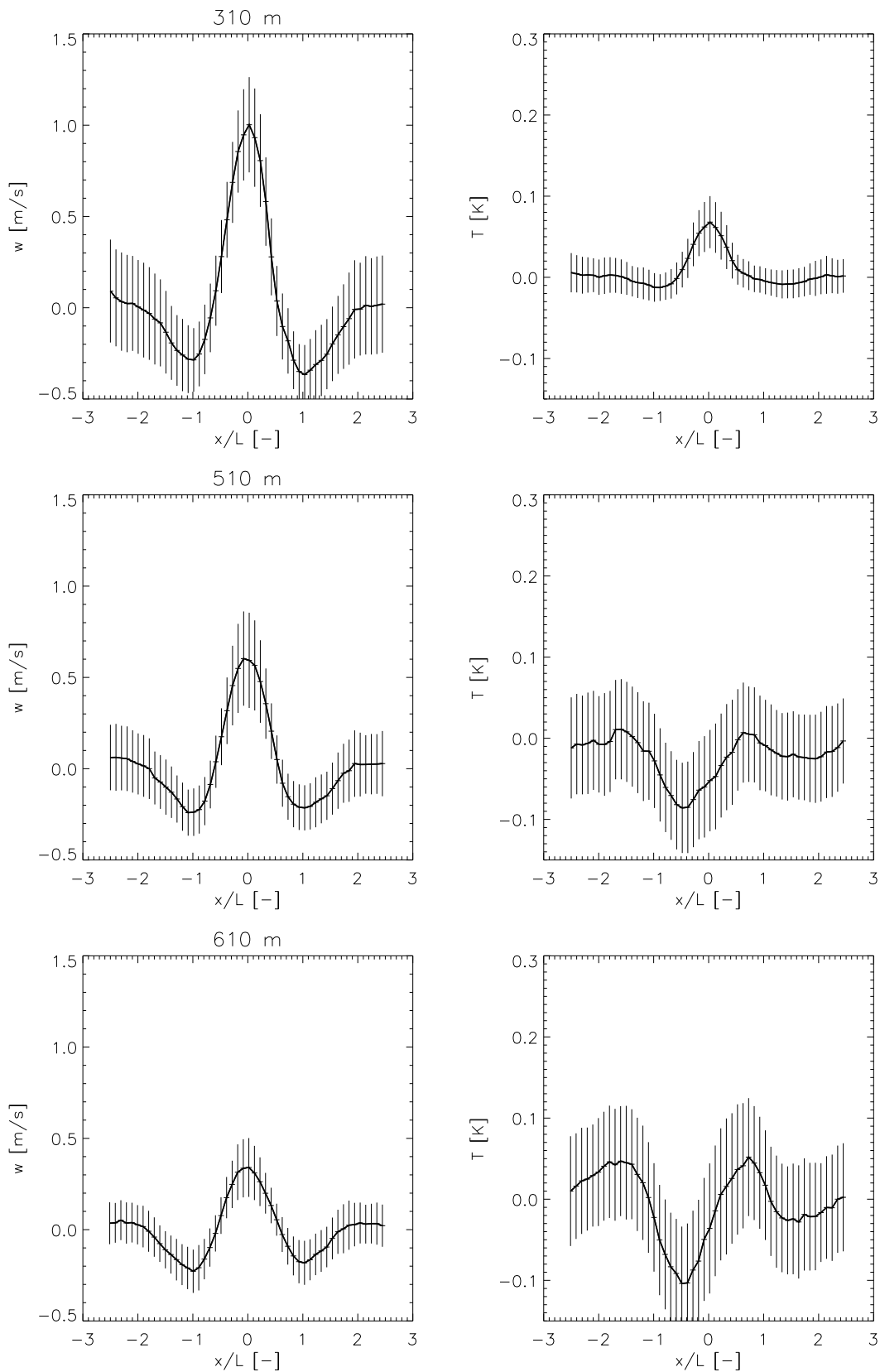
45

Figure 4.14: Part 2 of figure 4.13, showing how the shapes continue to evolve as height increases. The inversion height is not very strong for this figure, and starts between 200 and 300 meter.
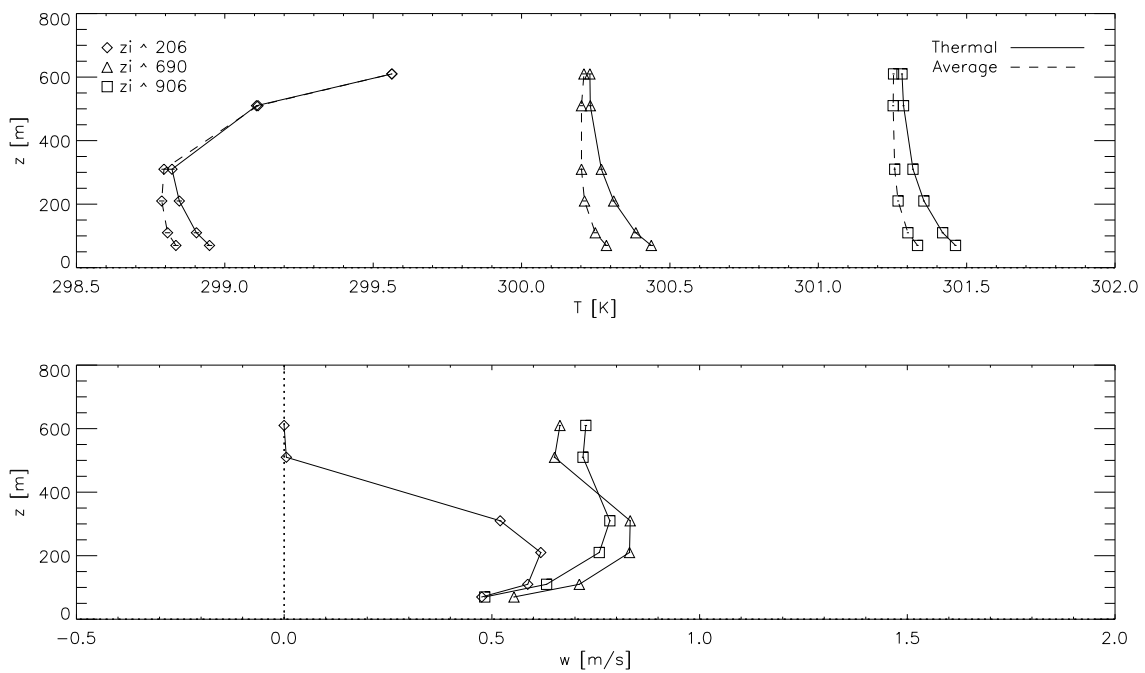
Figure 4.15: Figure illustrating properties and behaviour of thermals as they propagate higher into the convective boundary layer. The figure is based on three periods in a single LES run, from 1-2 hr, 5-6hr and 9-10hr simulated time, such as to visualize the dependence on inversion height. The upper panel shows average and thermal temperature, the lower panel shows the vertical wind speed of thermals.
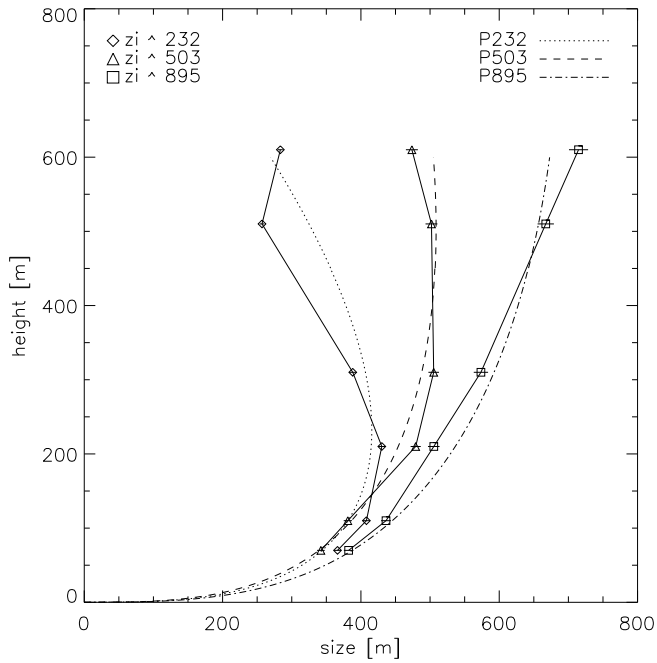
Figure 4.16: Figure illustrating thermal sizes with increasing height. The figure is based on the same three periods in a single LES run as before, 1-2hr, 5-6hr and 9-10hr simulated time. The solid lines show the behaviour of the LES periods, while the dotted, dashed and dash-dotted lines indicate the qualitative behaviour of Lenschow's prediction, equation 4.3, although the equation was rescaled with different factors to match the results.
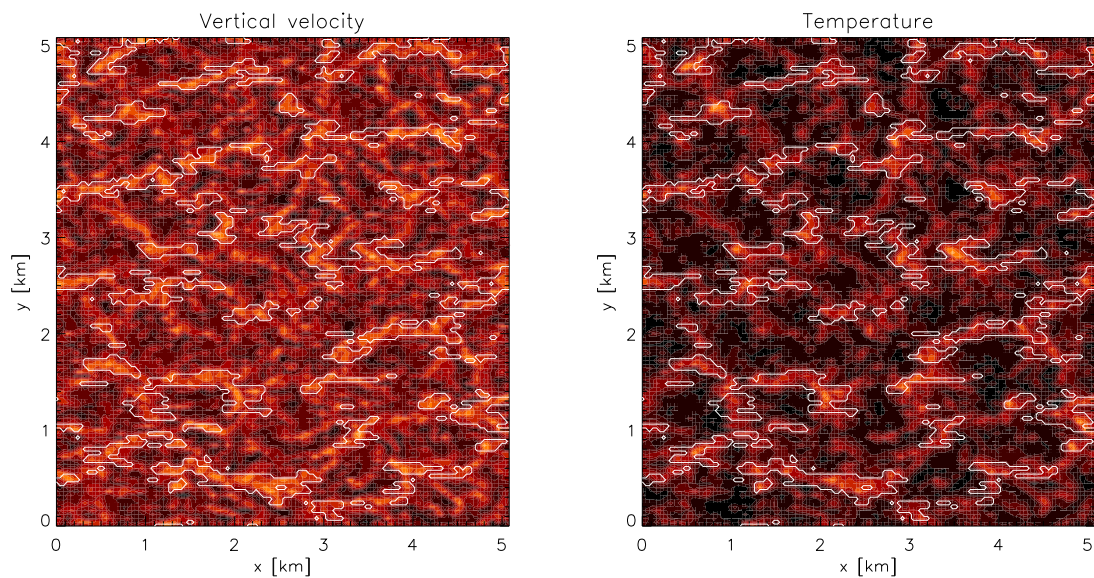

Figure 4.17: An instantaneous cross-section is shown of the vertical velocity and the temperature field, approximately 1.5 hours after startup, at 70m high. The white lines contour the parts of the cross section which have been identified as thermal, based on the time series at those points.

tubes and come in a wide variety of sizes and shapes.

To enable a quick comparison with the measurement series, temperature and vertical velocity pdfs are shown in figure 4.18. The lines are equal to those in figure 4.5; solid lines for the full pdf, red for the thermal and dotted for the non-thermal part. All data is collected from a single LES run from 1 to 2.5 hours simulated time, of which out of the 128x128 horizontal grid points, 100 random time series are chosen. Note that the qualitative behaviour of the LES dataset is much like the measurement series, including the area covered by the detected thermals. Due to the fact that these time series are more equivalent to each other than the measurement series, less characteristics are lost when making a pdf of the complete dataset. The skewness is much better visible, illustrating the role thermal structures play in this skewness, as the dotted non-thermal density function clearly shows a more symmetrical structure.

The probability density functions of the thermal itself have very Gaussian-like shapes, which is in accordance with the time series from Cabauw as well.

Also, effort has been made to reproduce the probability density functions of angle and size as figures 4.7 and 4.12(a) on the basis of LES results. Therefore, angle and size are calculated in the same way as was done for Cabauw data, using Taylor's hypothesis to transform the time series to a spatial system. Two LES runs were used as a basis for these calculations: one with a geostrophical wind forcing of 2 m/s and one with 4 m/s. Periods of 3-4 hours simulated time were used of both

Figures 4.19 shows the results. Note that while two different wind speeds were used to give a certain variety in wind speed, it does not nearly cover the broad range of wind speeds found in measurement data. This is especially notable in figure 4.19(b) where wavelet analysis descretizes the possible detected timescales of thermals. To cover this, exponentially increasing bin sizes where used. Note that this results in a figure qualitative much alike 4.12(a). For a quantitative analysis, a broader range of wind speeds would be needed. It would also be interesting to consider the size in the horizontal twodimensional spatial domain as is supplied by LES data. A smart algorithm would be needed however, to define where one thermal begins, and the other ends (reconsider figure 4.17).

Both measurements and LES data point at an ever-decreasing probability for increasing thermal sizes. The increasing probability and peaking at small sizes which can be seen well, especially in the LES results, is expected to be an artifact of the fact that a minimum size is needed for detection, and not true behaviour. The minimum scale which can be resolved by the LES calculations might also influence this finding in LES results, as the first peakings occur at 100 200m, which accounts to only a few 40m grid boxes.

Figure 4.19(a) shows behaviour which also shows some similarities with measured Cabauw data, yet it shows clear differences as well. While the figure is also somewhat symmetrical around zero and shows peaking at high positive and negative angles, it finds zero cases of angles in between. This cannot be explained by the small range in geostrophic wind, as no structural dependence on wind has been found. It might indicate that the peaks at highly tilted thermals are in fact artifacts of the reconstruction of the angle on the basis of detected thermals.

This could be resolved by determining the angle spatially, by looking at vertical instead of horizontal cross sections. This would give much more spatial information on the shape and
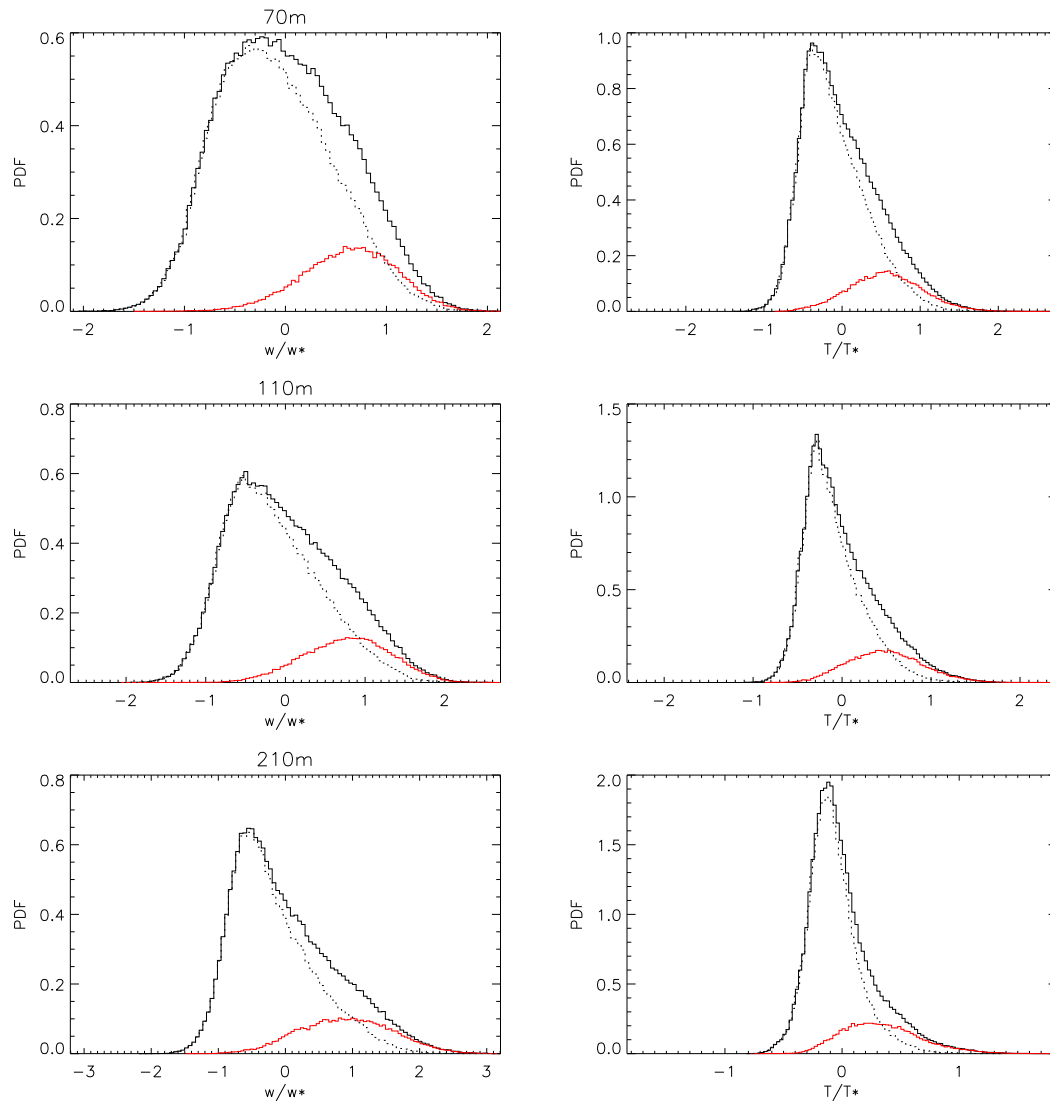
Figure 4.18: Probability density function of the characteristic properties of the the thermal structures based on 100 random time series of a LES run from 1 to 2.5 hours simulated time. Solid black lines represent the complete PDF, while the red and dotted lines represent the thermal and non-thermal parts of the PDF, respectively.

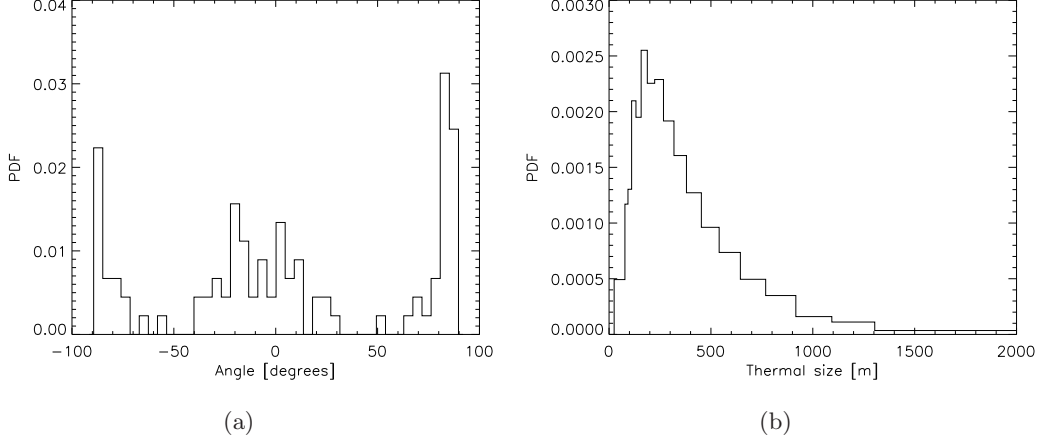structure of thermals as a function of height and is certainly an interesting subject for future research.

Figure 4.19: Distributions of thermal characteristics based on LES data. a) The distribution of angles with respect to the vertical and b) the probability density function of thermal size.

4.2   Mass flux approach

The flux contribution of the thermal structures in the CBL can be approximated by the use of the *Mass flux* approach (Siebesma et al, 2006; Neggers et al, 2009). Using the identification methods so far, this method can be evaluated step by step. Several assumptions are made, and the effects of each of these assumptions can thus be investigated.

The mass flux approach as applied to thermal structures is descibed in Neggers et al, 2009. Considering a parameter of interest φ, then the average value of such a parameter within the thermal is given by $\varphi_t = 1/a_t \int_{A_t} \varphi dA$, where the subscript t stands for thermal and the area A_t for thermal area. It follows that $\varphi_e = 1/(1 - A_t) \int_{1-A_t} \varphi dA$, where the subscript e denotes the enviromental value. Hence it is easy to see that $\varphi = a_t \varphi_t + (1 - a_t)\varphi_e$, with a_t the fractional area A_t/A. Applying this to the vertical velocity w and the potential temperature θ, the turbulent flux can be decomposed as (Siebesma, 2007)

$$\overline{w'\theta'} = a_t.\overline{w''\theta''}^t + (1 - a_t).\overline{w''\theta''}^e + M(\theta_t - \theta_e) \tag{4.6}$$

Where $M = a_t w_t$ the mass flux, and the overbar-t ($\overline{\varphi}^t$) represents averages over the thermal part of the data. In this text, the single primes indicate deviations from the complete average, while the double primes indicate deviations from the specific average, as given by:

$$\overline{w''\theta''}^t = \frac{1}{A_t} \int_{A_t} (w - w_t)(\theta - \theta_t)dA \tag{4.7}$$

The first and last term of the right hand side of equation 4.6 together represent the flux in the thermal structures, while the second term represents the flux in the environment. Often now

the first term is neglected, which is justified if the subthermal fluctuations are uncorrelated. The result is

$$\overline{w'\theta'} = (1 - a_t).\overline{w"\theta"}^e + M(\theta_t - \theta_e) \tag{4.8}$$

Which is used in, for example, Neggers et al, 2009. The first term can then be modelled by the well known K-diffusion methods, while the second term is modelled by the mass flux approximation. The further approximations involved are the estimation of the values of w_t, θ_t and a_t. It is common to estimate a_t as a constant. Then the spread of w and θ is approximated as being normally distributed, with w_t and θ_t being the top a_t percent of the gausssian probability function. It can then be shown (Appendix B) that:

$$
\begin{aligned}
w_t &= \sigma_w D(a_t) \\
\theta_t &= \overline{\theta} + \sigma_\theta D(a_t)
\end{aligned}
\tag{4.9}
$$

Where D a function which is commonly tabulated. While averages are known, the standard-deviations have to be approximated by similarity relations (based on Siebesma et al, 2006; Neggers et al, 2009)

$$
\begin{aligned}
\sigma_w &= 1.3 \left[u_*^3 + 0.6 \frac{z}{z_i} w_*^3 \right]^{1/3} \left(1 - \frac{z}{z_i} \right)^{1/2} \\
\sigma_\theta &= b \frac{\overline{w'\theta'}_h}{\sigma_w}
\end{aligned}
\tag{4.10}
$$

Where the subscript h indicates the flux at the given height. Note that the the calculation of σ_θ uses the calculated values of σ_w.

Figure 4.20 evaluates the effectiveness of these relationships. The figure is based on four 1-hour LES-periods. Two periods feature a surface flux of 0.06 Km/s, one period is taken from 1-2 hours simulated time, the second from 4-5 hours simulated time, yielding a large increase in inversion height. The other two periods are taken from 3-4 hours simulated time with a surface flux of 0.03 Km/s and 0.09 Km/s.

Within the given periods, all 128x128 time series are analysed to determine the thermal locations as before. Then, relevant parameters are calculated each minute, on the basis of a cross section at 70m high. The horizontal axis of figure 4.20 denotes the true value of σ, while the vertical axis shows the quality of relationships 4.10.

On the basis of the least squares principle, the coefficient b used in the calculation of σ_θ is found to be $b = 1.215$.

Figure 4.20 shows that the behaviour of σ_w is somewhat less accurate for the individual LES simulations, but as will be seen later, the average quality of is good enough for the mass flux approach.

This then completes the approximation of parameters within the mass flux approximation. Hence, in order to evaluate the mass flux approximation, the relationship

$$a_t.\overline{w"\theta"}^t + M(\theta_t - \theta_e) = [M]([\theta_t] - [\theta_e]) \tag{4.11}$$

has to be evaluated. In equation 4.11, square brackets ([,]) indicate that the used values are the results of parameterisations.

The terms on the right hand side can be directly measured using the thermal identification
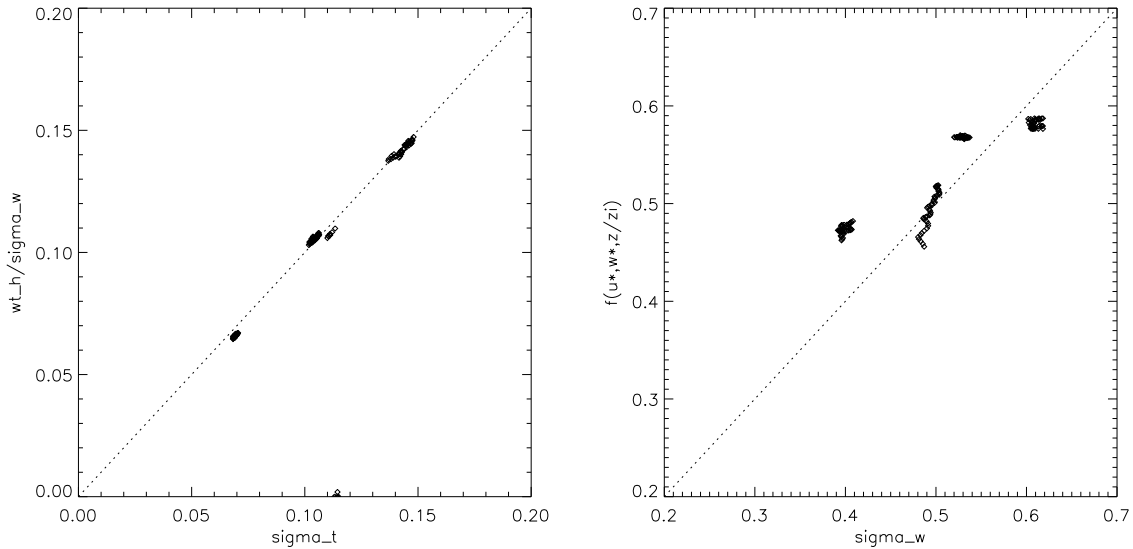
Figure 4.20: Equations 4.10 are evaluated based on four 1h periods in LES data with varying inversion height and surface flux.

system, and are plotted on the x axis of figure 4.21(a), based on cross-sections of LES data. This figure demonstrates the effect of the above approximations in the following order:

*1)*The first term on the left hand side is neglected.

2) The fraction $[a_t]$ is now assumed constant and to be equal to the average of the measured $[a_t]$. $[w_t]$, $[\theta_t]$ and $[\theta_e]$ are estimated by assuming a normal distribution and using equations 4.9. However σ_w and σ_θ are directly measured.

3) Now $[\sigma_w]$ and $[\sigma_\theta]$ are parameterized according to equation 4.10.

4) Lastly, $[a_t]$ is now customly adjusted such that the parameterisation best fits the data. The least squares method is used to find $[a_t] = 9.9\%$.

It is interesting to see how the parameterisation exactly works and is dependent on the choice of $[a_t]$. The first approximation obviously underestimates the flux as a term is neglected. This loss is perhaps larger than expected, though. This underestimation is overcompensated by approximating a Gaussian distribution shape in the second step. Assuming all values to be in the very tail of the distribution, while still using the same surface fraction gives a large over-estimation of the flux. This could be expected, considering that figure 4.18 shows a Gaussian like shape for the distribution of the thermal part of the probability density functions, which results in a large amount of points which do not reside in the tail of the total PDF.

The third approximation differs very little from the second, indicating that the parameterisation of the standarddeviations is very effective within this application. The result is then corrected by taking a surface fraction such that the mass flux approximation is effective. This will put thermal properties further in the tail, somewhat increasing $[t_t]$ and $[w_t]$, but the decrease of $[a_t]$ has a larger influence.

With the commonly chosen value of $a_t \approx 10\%$ the mass flux approach seems quite able to capture the average behaviour of flux in a given period, although the variations within each period are less well followed. This is mostly caused by the approximation of a constant a_t.
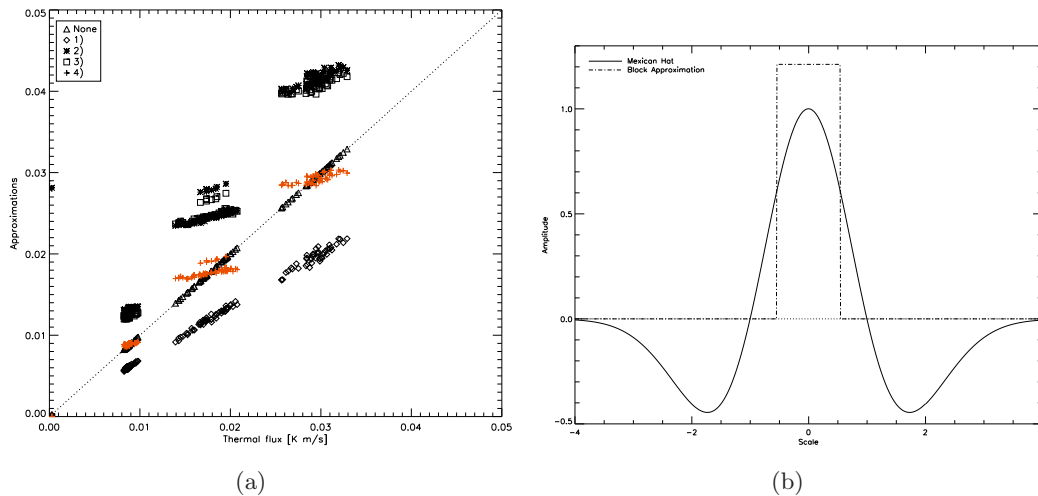
(a)   (b)

Figure 4.21: The mass flux approximation. a) The left hand side of equation 4.11 is plotted on the x axis, while the vertical axis shows the quality of a cascade of approximations, of which the final step is shown in red. The numbers in the legend correspond to the point at which the cascade of approximations is applied. b) Shows the principle of these approximations, with the Mexican Hat function representing the thermal (the thermal part is defined ad the positive part, which ranges from -1 to 1), and the block function representing the mass flux approximation.

The dependence on a_t could have been expected, as the mass flux approach is essentially a block function approximation, in which the smooth varying shape of a thermal has to be estimated using a block function, which hence has to have a smaller area to produce the same amount of transport. Figure 4.21(b) demonstrates this principle by plotting the Mexican Hat function as an estimation to a thermal, and the estimation using a block function with the same relative a_t and w_t to the Mexican Hat as used in the final step of equation 4.21(a). Indeed, using this values it is found that the area under the the thermal part of the Mexican Hat is equal to around 92% of the area under the block function.

Before studying the effects of variation of a_t deeper, the question is whether the above results can be found in Cabauw measurements as well. The problem in that, however, is that figure 4.21(a) is based on horizontal cross sections, while Cabauw measurements only consider the time dimension. To study what the effect is of a transformation to the time dimension, 100 randomnly chosen time series within the simulation results used for figure 4.21(a) are analyzed in the same way as the cross sections.

Figure 4.22(a) shows the result of this time-based analysis. It shows that where the simulations with different surface flux were easily separable in figure 4.21(a), no such separation is visible anymore in the time series. The random sampling error induced by considering a random time series is thus much larger than the differences caused by changing the driving force. However, even with this random sampling error, the same qualitative behaviour is found, where the neglect of sub-thermal transport causes underestimation of the flux, and the modelling of a thermal by the tail of a normal distribution overcompensates this effect.

The value of $[a_t]$ in the last approximation is set to 10% to match the results from figure 4.21(a). Indeed it is witnessed that this value leads to a good approximation in figure 4.22(a) as well.

Figure 4.22(a) has demonstrated that results from time series show a much larger spread
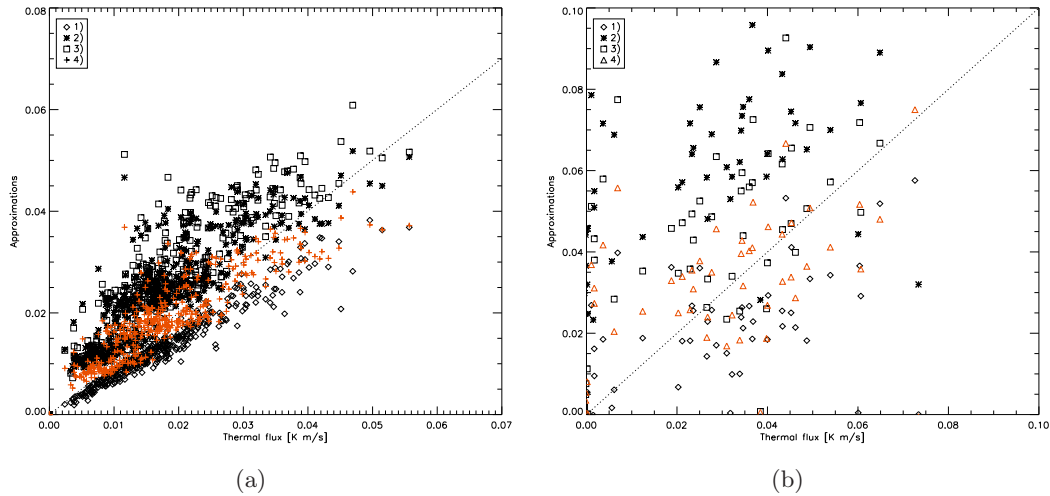
(a)   (b)

Figure 4.22: The mass flux approximation evaluated on the basis of time series. a) 100 ramdonly chosen time series from the same LES simulations as used for figure 4.21(a) are analyzed to directly show the result of changing dimensions to time. b) Shows the analysis performed on the basis of 24 days of Cabauw time measurement with three 1.5hr periods used per day. The numbers in the legend correspond to the point at which the cascade of approximations is applied.

in variables due to random sampling errors which are nonexistant in horizontal cross section analysis. However, when accepting this spread it is still possible to evaluate the mass flux approximation on the basis of time series. This is comforting, as random sampling errors are inevitable in point measurements.

Figure 4.22(b) then shows exactly the same steps, but now applied to time series data from Cabauw. The 24-day dataset of which inversion height is available is used again, with three 1.5hr from 10.30h to 15.00h periods used per day. The spread is now much larger than the LES time series, which is expected as the LES introduces a random sampling error based on the same dataset, while Cabauw data is based on days with differing conditions. Yet, the same qualitative behaviour is witnessed, and the red triangles representing the final results are centered around the 1:1 line, indicating that while the spread is large, the average behaviour confirms the conclusions which were made based on LES data.

To study how results are dependent on the surface fraction a_t, which varies as stricter or less strict definitions of thermals are used, an algorithm is used to expand (include the area surrounding the thermal) or contract (remove boundary areas) the detected thermal fields . This variation is done on the basis of horizontal cross sections.

This is done by representing the information on which part of the cross section is thermal, and which is not, as a seperate cross section layer with only bits. Ones represent points in the cross section which belong to a thermal, zeros the points which do not. Now a smoothing filter is applied, with the result that a zero-to-one jump now is a smooth transition from zero to one over a range of points. Varying the treshold of which points are now labeled as thermal allows us to expand or contract the thermal field.

The mass flux and sub-thermal flux as a function of the area fraction are shown in figure 4.23, normalized by the area which has been detected by wavelet analysis. This also gives

better insight into what the algorithm actually detects.

  The figure is based on the average of 60 cross sections within the 60 W/m^2, $U_{geo} = 4$ m/s run, from 9-10hr, and clearly shows the dependency of fluxes on the definition of the edges of thermals. As expected, the sub-thermal flux practically linearly increases with area, since this represents local turbulent transport which is expected to be roughly constant throughout the area. For low a_t, the mass flux increases much faster, which is expected as this now encloses more and more of the strong non local transport for which the thermals are responsible. As a_t increases, the mass flux levels off and even decreases. This is expected as well, as at this point the full thermal transport is enclosed, and further increases encloses environmental negative flux, hence decreasing the averages w_t and t_t. This further confirms the fact that cold descending air surrounds the warm and rising thermals. This phenomenon might also explain why the subthermal turbulent fluxes suddenly drop for large relative surface fractions. At it's maximum, the contribution from the mass flux is typically between 30 and 40 percent of the total flux.

Some of the noisiness of both curves can be attributed to the algorithm which expands the area instead of physical phenomena.

The cross section a_t detected by the wavelet algorithm detects an area fraction around 18% in all cases, which can be seen to be located at the point where the mass flux levels off. In fact, it is typically close to the point where the difference between the mass flux and the sub thermal flux is maximal, which is expected to be related to the fact that the local wavelet power spectrum detects the largest power using this boundaries.
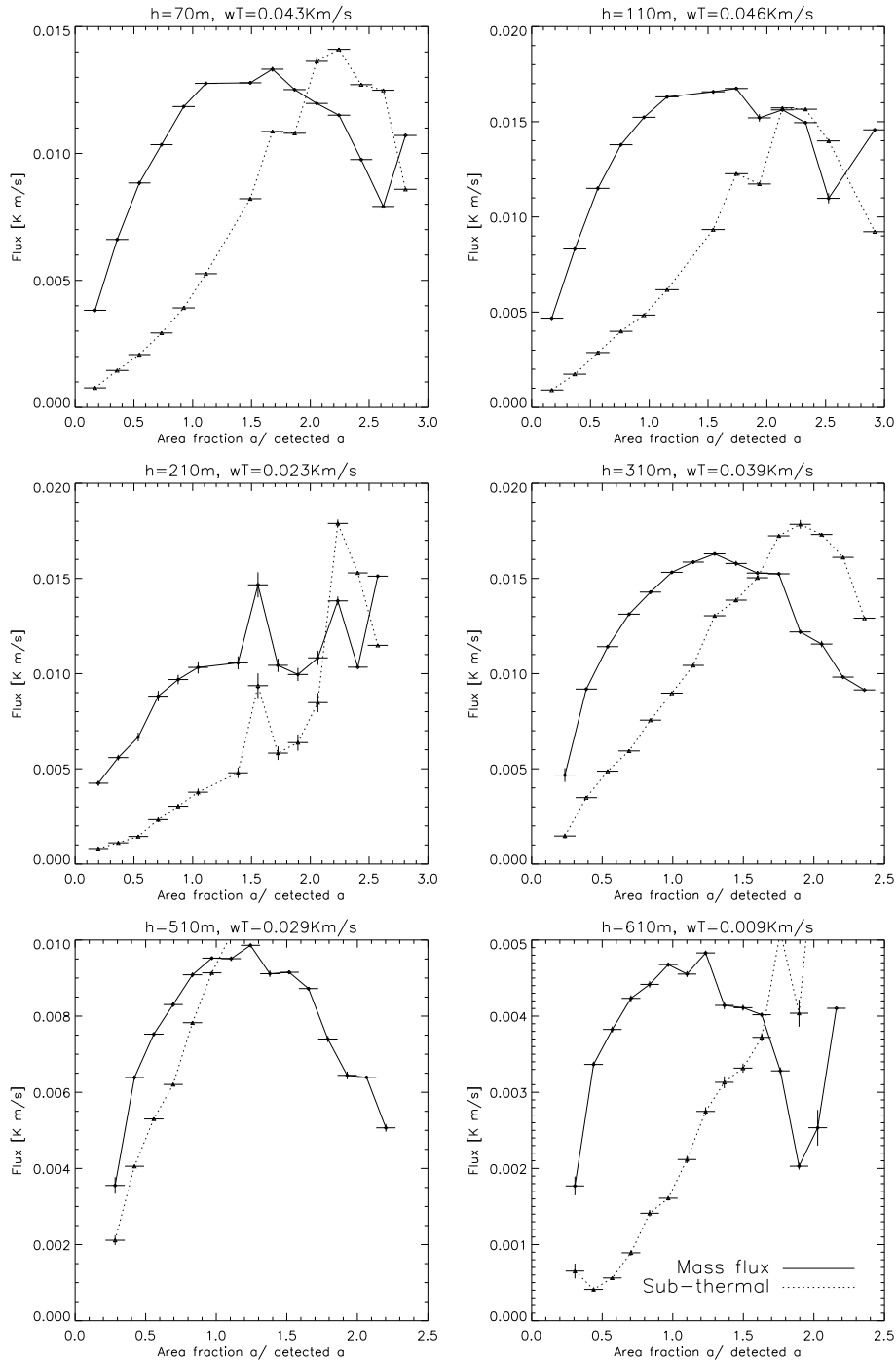
Figure 4.23: The area which labels thermals is expanded and contracted, visualising how the mass flux and subthermal flux depend on the definition of a thermal. The average total flux is mentioned in the title.

Chapter 5

Conclusion

In this report, timescales of turbulent transport in the vertical boundary layer have been investigated, with the focus on heat transport. A dominant timescale has been defined as the inverse of the integral of the cospectrum S_{wT} of vertical velocity and temperature with the frequency as weight, normalized by the integrated cospectrum itself.

Modelling the shape of the cospectrum by a simplified curve, the peak of fS_{wT} is found to be located at four times this dominant timescale. A loss function has been derived based on this cospectrum, which predicts the loss in Eddy Covariance point measurement methods due to finite averaging time as a function of averaging time and the dominant timescale.

Eddy Covariance measurements in Cabauw, on days selected for their convective and cloudless behaviour, showed behaviour very similar to this prediction, confirming that the loss is indeed predictable. In order to predict this loss, knowledge of the dominant timescale is required. An improved empirical equation is proposed to determine this scale as a function of height, Obukhov length and wind speed. No evidence is found for the presence of unpredicted systematic flux on very large timescales.

LES results confirmed the qualitative behaviour of the loss as a function of averaging time, but showed some deviation for large averaging times.

The method of multiresolution decomposition, used to find a gap scale which distinguishes turbulent motions from mesoscale motions is also evaluated. It is found that the dominant timescale of transport closely approaches the gap scale in unstable situations, such that turbulent and mesoscale motions are no longer seperable.

Secondly, thermals have been studied as they form structures responsible for large contributions to turbulent transport. Two algorithms are introduced, which together have been found able to detect thermals as they pass an observation site. The first algorithm scans for vertically coherent deviations in vertical velocity and temperature, while the second algorithm improves the find by performing wavelet analysis to distinguish the exact edges of the thermals.

Timeseries from Cabauw were supplemented with fields resulting from LES simulations to do a structural analysis of thermals. Thermals detected in LES time series were found to behave similar to measurements in Cabauw.

It is found that thermal structures appear in time series in a Mexican Hat-like form, including portions of descending, cold air before and after the thermal. Thermals appear in all sizes, with probability decreasing with size, and the mean size independent of wind speed. Depen-

dence has been found on height, as thermals grow all the way up to the inversion height.

The air in a thermal accelerates from the ground up to the inversion, driven by a temperature excess which turns into a deficiency above the inversion, decelerating the air.

The average flux within thermals is found to be well approximated using the mass flux approximation, when it used on the basis of an updraft area of 10 percent. This non-local transport can typically represent 40 percent of the total vertical turbulent transport of heat.

Appendix A

Probability Density Function of Thermal size

One of the problems in measuring the size of a thermal is that the distribution of measured sizes can be partially explained by the fact that EC covariance measurements can be seen as taking a one-dimensional cross section of space. This means that, assuming a certain shape of the thermal, when the thermal passes over the measurement tower, the tower would typically not cross the thermal exactly through the center. The result is that, even when all thermals would have exactly the same size, still a wide range of thermal sizes would be measured. To be able to nevertheless make statements on the measured data, lets take a look at the probability density function which is introduced by the phenomenon described above.

For example, assume a thermal has the shape of a circular column. At any height, this would account to a circular cross section with a certain radius R.

Now the probability of crossing the thermal with an offset from the center x (Figure A.1) is the same for all x. Hence the probability function of x would become:

$$f_x(x) = \begin{cases} \frac{1}{2R} & -R < x < R, \\ 0 & otherwise \end{cases} \tag{A.1}$$
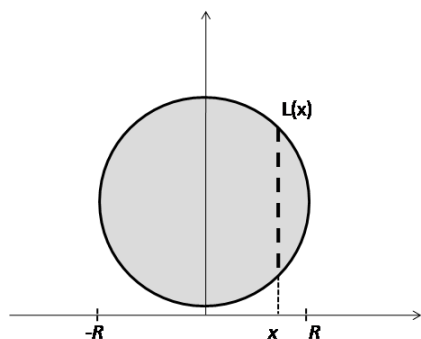

Figure A.1: Figure illustrating how a distribution of measured thermal sizes can occur through the offset x with which thermals pass the measurement tower.

As the thermal cross-section is circular, crossing the thermal starting at coördinate x will mean a crossing length $L(x)$ of:

$$L(x) = 2\sqrt{R^2 - x^2} \tag{A.2}$$

This gives us immediately the interesting result that the expected value for the mean measured L is *not* the same as the diameter $D = 2R$:

$$E\left[L(x)\right] = \int_{-\infty}^{\infty} L(x) f_x(x) dx = \frac{\pi}{2} R \tag{A.3}$$

The cause of this is of course that the theoretic maximum path length L in the thermal is $2D$, and the minimum is 0. In reality, only thermals with a path length L larger than a certain treshold will be detected.

Now lets take a look at the expected PDF. The PDF of L, $f_L(L)$ can be calculated by using the invariance of a differential area, or $|f_x(x)dx| = |f_L(L)dL|$. Using equation A.2 to find dL/dx, the probability density function becomes:

$$f_L(L) = \begin{cases} \frac{L}{4R} \frac{1}{\sqrt{R^2 - L^2/4}} & 0 < L < 2R, \\ 0 & otherwise \end{cases} \tag{A.4}$$

Figure A.2 shows the behaviour of this probability density function. Note how the probability increases towards larger scales up to the maximum scale.
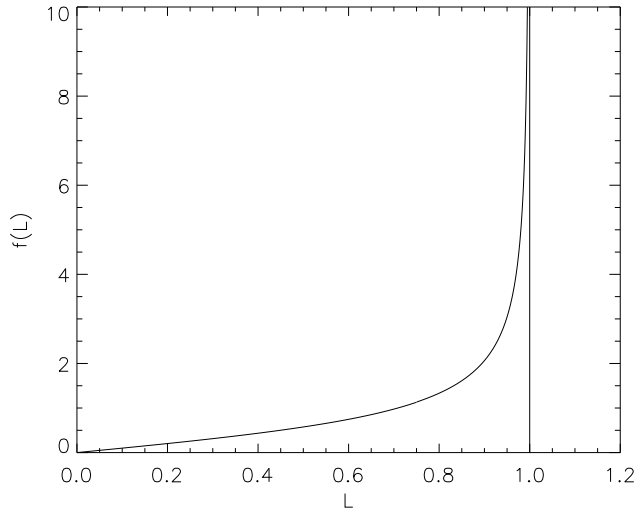

Figure A.2: The behaviour of equation A.4.

Appendix B

Percentile of a normal distribution

Given a Gaussian probability distribution of a variable φ, we are interested in the mean $\overline{\varphi}_a$ of the top fraction a of the Gaussian function. Assuming zero mean for simplicity, we have:

$$\overline{\varphi}_a = \int_x^\infty \frac{\varphi}{\sigma_\varphi \sqrt{2\pi}} \exp\left[-\varphi^2/2\sigma_\varphi^2\right] \tag{B.1}$$

Where x is the coordinate for which the top fraction a of the Gaussian starts, hence:

$$a = \int_x^\infty \frac{1}{\sigma_\varphi \sqrt{2\pi}} \exp\left[-\varphi^2/2\sigma_\varphi^2\right] \tag{B.2}$$

Performing the integrations, the last equation can be inverted to express x as a function of the inverse error function of $1 - 2a$, $erf^{-1}(1 - 2a)$. This can be inserted into equation B.1, after which integration leads to the conclusion that $\overline{\varphi}_a/\sigma_\varphi$ is a function of a only.

$$\begin{aligned} \overline{\varphi}_a &= \sigma_\varphi \sqrt{\frac{2}{\pi}} \left[e^{-erf^{-1}(1-2a)} - 1\right] \\ &\equiv \sigma_\varphi D(a) \end{aligned} \tag{B.3}$$

With erf^{-1} the inverse of the error function. Given a variable ϕ with a nonzero mean μ, a coordinate transformation easily yields the same results for $\varphi = \phi - \mu$. The values of $D(a)$ are tabulated in Neggers et al (2008).

Appendix C

Correcting t^* for high-frequency losses in LES results

When comparing values of t^* from LES results with t^* found from high frequency measurement sites like the sonic anemometers in Cabauw, a significant difference is found in typical order. The definition of t^* places more weight on high frequency values of the spectra, which is especially done since it is well known (Kaimal, 1972) that the high frequency part of the spectrum behaves better than the lower frequency part. Adding to this effect is the fact that numerical schemes for calculating the spectra introduce larger uncertainties for the low frequency part. However, LES results model the high frequency turbulence in subgrid schemes, and this part will therefore not be found in a spectrum calculated based on a LES simulated time series. This results in an overestimation of t^* when implementing the same techniques on LES results as were used on measured data.

In this appendix, a correction for this effect will be sought using the same assumption for the shape of the spectrum as was used in chapter 3. For convenience, the definition of timescale t^* (equation 3.1 and the Kaimal (equation 3.11) are repeated here:

$$t^* = \left[\frac{\int_{-\infty}^{\infty} f S_{wT}(f) df}{\int_{-\infty}^{\infty} S_{wT}(f) df} \right]^{-1} \tag{C.1}$$

$$S_{wT}^f(f) = \frac{4 F_T t^*}{\left(1 + \frac{4 f t^*}{p-1}\right)^p} \tag{C.2}$$

Using the assumption of the spectral shape, it is clear that inserting this equation into the equation for t^* yields $t^* = t^*$ as required. However, in the LES results, assume no contribution can be found on frequencies higher than a certain cutoff frequency f_c, which is the lowest of either the timestep or the $\delta x/U$. Assuming the LES resolves the important part of its turbulent flux, it is assumed that integrating the spectrum still yields F_T, the flux. Hence, an uncorrected timescale, which will be denoted by t', is found characteristic of the LES timeseries:

$$\frac{1}{t'} = \int_0^{f_c} \frac{4 f t^*}{\left(1 + \frac{4 f t^*}{p-1}\right)^p} df \tag{C.3}$$

Using coördinate transformations as before and inserting $p = 7/3$ for the heat flux cospectrum, this equation can be directly integrated to find:

$$\frac{1}{t'} = \frac{1}{t^*}\left[\frac{1}{3}x^{4/3} - \frac{4}{3}x^{1/3} + 1\right] \tag{C.4}$$

Where x is a coördinate related to the cutoff frequency and dominant timescale: $x = 1/(1 + 3f_c t^*)$. No LES timeseries at very small heights (< 10 m) will be considered, since a very high resolution model is needed to give a correct representation for this. Therefore, t^* will be in the order of 50-100 seconds (see figure 3.3). For the most extensively used LES run, a horizontal resolution of 40 m is combined with 4 m/s wind speed, hence the cutoff frequency will be of the order of 10 seconds. Thus, $x \approx 0.01$ and hence the first term in brackets can be neglected. The resulting equation is:

$$\frac{t^*}{t'} = 1 - \frac{4}{3}x(t^*)^{1/3} \tag{C.5}$$

The Newton-Raphson method will be used to find a t^* such that this equation is satisfied for the given t'.

Bibliography

[1] F. Bosveld. The Garderen experiment, micro-meteorological observations. KNMI-WR-99-03. 1988-1989: Corrections KNMI-TR-276.

[2] M. de Haij. Evaluation of a new trigger function for cumulus convection. 2005.

[3] F. Dekking, C. Kraaikamp, and H. Lopuhaä. *A modern introduction to probability and statistics*. Springer, 2005.

[4] J. Kaimal, J. Wyngaard, Y. Izumi, and O. Coté. Spectral characteristics of surface-layer turbulence. *Quart. J. R. Met. Soc.*, 98:563–589, 1972.

[5] M. Kanda, A. Inagaki, M. Letzel, S. Raasch, and T. Watanabe. LES study of the energy imbalance problem with Eddy Covariance fluxes. *Bound. Layer Meteorology*, 110:381–404, 2003.

[6] D. Lenschow and P. Stephens. The role of thermals in the convective boundary layer. *Bound. Layer Meteorology*, 19:509–532, 1980.

[7] R. Neggers, M. Köhler, and A. Beljaars. A Dual Mass Framework for boundary layer convection. part i: Transport. *J. o.t. Atmos. Sciences*, 66:1465–1487, 2008.

[8] J. Schalkwijk and H. Jonker. LES study on the flux imbalance in the convective boundary layer. 2008.

[9] A. Siebesma, P. Soares, and J. Teixeira. A combined eddy-diffusivity mass-flux approach for the convective boundary layer. *J. o.t. Atmos. Sciences*, 64:1230–1248, 2006.

[10] R. B. Stull. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 1988.

[11] D. Vickers and L. Mahrt. The cospectral gap and turbulent flux calculations. *J. Atmos. and Oceanic Technol.*, 20:660–672, 2002.
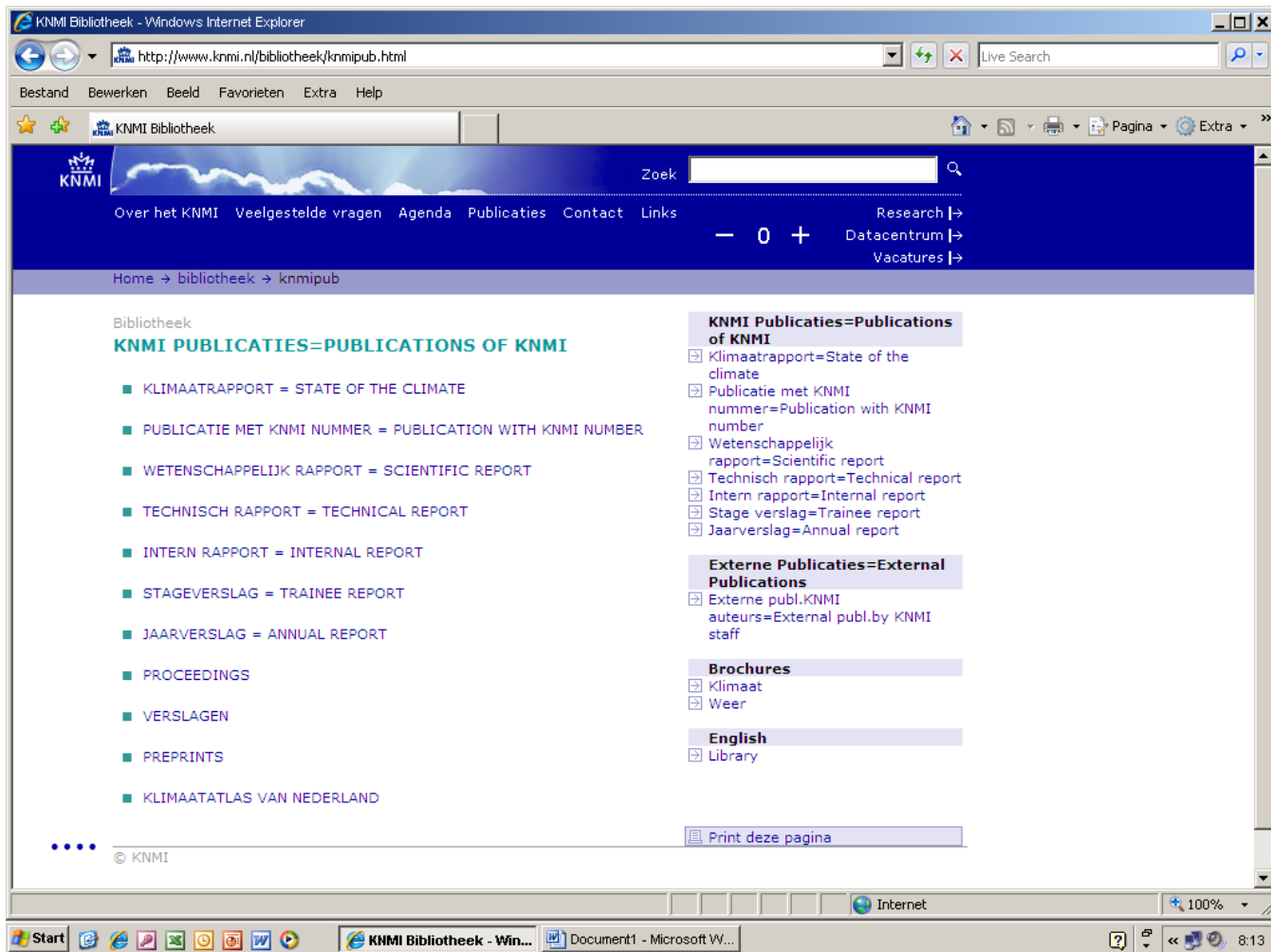
All titles of KNMI-publications (and a full text PDF for the most recent ones) can be found on

http://www.knmi.nl/bibliotheek/knmipub.html

If you have any questions, please contact us: bibliotheek@knmi.nl