

Characteristics of Convective Turbulence in the Surface Layer Investigated by Principal Component Analysis

E. P. WEIJERS

Department of Meteorology, Vrije Universiteit, Amsterdam, the Netherlands

A. VAN DELDEN

Institute for Marine and Atmospheric Research, Utrecht University, Utrecht, the Netherlands

H. F. VUGTS AND A. G. C. A. MEESTERS

Department of Meteorology, Vrije Universiteit, Amsterdam, the Netherlands

(Manuscript received 25 February 1994, in final form 1 June 1994)

ABSTRACT

Principal component analysis (PCA) was applied to 182 half-hour runs containing time series of turbulent wind velocity and temperature measured in the convective atmospheric surface layer. A field experiment with four sonic anemometers on the vertices and one in the centroid of a square (with sides of 80 m) was performed to obtain the necessary dataset. Physical explanations of the most important eigenvectors are presented. Two of the major principal components (PCs) identify the variance in wind speed along and across the background wind direction. Always, one major PC accounts for the presence of large-scale thermal activity: periods with higher (lower) temperatures coincide with lower (higher) wind speeds, convergence (divergence) in the wind fields, and upward (downward) movements. As an application, variance in the velocity fields was expressed in terms of horizontal divergence and vertical vorticity. These can be derived directly from the eigenvectors when PCA is combined with a planimetric method. Using the PC that identifies thermal activity, it is found that the magnitude of divergence increases and the magnitude of vorticity decreases when atmospheric conditions become more unstable. It is found that the (absolute) ratio between vorticity and divergence scales with a function of the friction velocity divided by the convective vertical scaling velocity. Both kinematic parameters are larger for updrafts than for downdrafts. It is concluded that PCA can be a useful tool to distinguish variance of thermal and nonthermal origin and in the estimation of the kinematics of dominant flow fields.

1. Introduction

In the convective atmospheric boundary layer, turbulence is dominated by discrete thermally driven updrafts and downdrafts (Young 1988). The updrafts transport relatively warm air and originate in the surface layer (SL), while the downdrafts contain cooler air from above the mixed layer. The "footprints" of these thermal structures, usually called "plumes," have diameters and depths on the order of the depth of the SL (~ 100 m). When probing the SL, they can be recognized as recurring ramplike signals in the temperature time series. Figure 1 displays such a temperature ramp, which was observed at a height of 31 m, with a mean background wind speed of 4.5 m s^{-1} . Charac-

teristic is the gradual increase in temperature with time followed by a sharp drop at the upwind edge ("micro-front"). The plumes are surrounded by periods with relatively quiescent cooler air. Many studies have been made on several of their properties (e.g., Kaimal and Businger 1970; Antonia et al. 1982; Schols et al. 1985). Here we will focus on the horizontal circulation fields induced by these thermal events. Little is known regarding these flow fields. Due to continuity, one expects the upward movements to be accompanied by a horizontally convergent airflow, while the downward movements will cause the underlying wind field to be divergent. Although it is possible to derive such information from measurements made at several points along one mast (Wilczak 1984) or with aircraft (Williams and Hacker 1992), this is not an easy task.

In the study of convective boundary layer features, conditional sampling is a popular approach (e.g., Khalsa 1980; Sikora and Young 1993). With this technique, data are sorted into thermal and nonthermal categories according to some chosen indicator function,

Corresponding author address: H. F. Vugts, Department of Meteorology, Vrije Universiteit, De Boelelaan 1085, 1081 HV Amsterdam, the Netherlands.
E-mail: vugh@geo.vu.nl

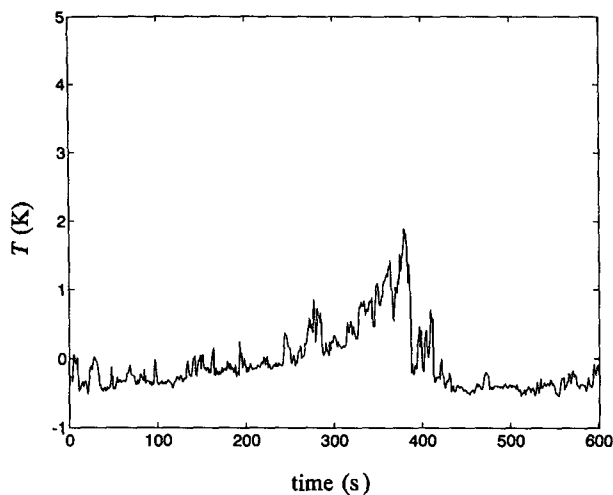


FIG. 1. Time series of temperature indicating the passage of a thermal structure in the SL.

Indicators of thermal events can be temperature, vertical velocity, or turbulence, sometimes combined with a requirement that the indicator variable maintain a certain value over a specified time duration (e.g., Weijers et al. 1994). Once the criteria have been established, conditional averages can be calculated for any set of variables. However, the weakness of this method is that different criteria may lead to different results (e.g., Godowitch 1986).

The incentive for this study was then twofold: 1) to estimate the kinematic properties of the horizontal wind fields that are related to thermal activity in the SL and 2) to find out whether the objective technique of principal component analysis (PCA) can be used in studies of atmospheric turbulence. Meteorologists often use PCA as a tool to investigate patterns of velocity, pressure, temperature, or precipitation over a large spatial area (e.g., Green et al. 1993). Here, we discuss how the method performs on a “micrometeorological scale” with typical time and space lengths of 1 min and 100 m, respectively. As far as we know, such an application has not been described in literature before.

The central idea of PCA is to reduce the dimensionality of the dataset in which there are interrelated variables while retaining as much as possible of the variation present in the dataset. This is achieved by transforming to a new set of variables, the “principal components” (PCs), that are uncorrelated and ordered so that the first few retain most of the variation. The PCs are defined by the eigenvectors of the correlation (or variance–covariance) matrix formed from the ensemble of meteorological observations. The eigenvalues give the variances of the PCs and are therefore a measure of their importance in explaining variation. Important properties of the technique are 1) the eigenvectors and corresponding PCs are derived by an ob-

jective mathematical procedure; 2) some of the eigenvectors can be identified with physically important patterns in the original data; and 3) the eigenvectors provide an orthonormal set of basis vectors in terms of which the original observations can be represented, or, by selection of some of the eigenvectors, approximated. Here, the use of PCA is motivated by the concept that the flow field can be regarded as a superposition of different flow patterns. After the application of PCA, the resulting eigenvectors are expected to reveal the dominant patterns, and the corresponding eigenvalues give their importance.

In this paper, results of a PCA on a dataset containing the turbulent perturbations of wind velocity and temperature are discussed. To obtain the data, a field experiment was carried out using an array of five sensors, spanning an 80 m × 80 m area in the SL. The required mathematical procedures are given. The major eigenvectors are interpreted. It is derived how the characteristics of a flow field can be expressed in terms of horizontal divergence and vertical vorticity by use of PCs. The dependence of these kinematic parameters on atmospheric stability is discussed.

2. Methods of analysis

We start with the description of principal component analysis. This is followed by an explanation of the mathematical procedure of estimating horizontal divergence and vertical vorticity of vector fields and how these kinematic variables can be derived directly from the PCs. Finally, the criteria used to select the PCs relevant for this study are formulated.

a. Principal component analysis

The variant of PCA employed here tends to isolate groups of observations with recurring similar spatial velocity patterns. The method itself is described extensively by Jolliffe (1986) and Preisendorfer (1988).

Consider the dataset as an ensemble of a large number of realizations of the stochastic vector $\mathbf{F} = (f_1, \dots, f_q)$. In this study, the components of \mathbf{F} represent the turbulent parts of q meteorological variables at some time. In our PCA computations we use standardized variables. These are defined as the components of a new stochastic vector \mathbf{U} : $U_i = f_i / \sigma_i$, $i = 1, \dots, q$, with σ_i the standard deviation of the i th variable. The elements of the correlation matrix \mathbf{S} are then given by $S_{ij} = \langle U_i U_j \rangle$, $i, j = 1, \dots, q$. The brackets stand for ensemble (or time) averaging and it is supposed that the mean of any variable is zero (which is, by definition, the case for turbulent components).

In some applications \mathbf{S} is a covariance–variance matrix. The reason that a matrix of correlation coefficients is preferred here is that, in case of large differences between the variances of the elements of \mathbf{F} , the

variables whose variances are largest will tend to dominate the first few PCs. Especially when the variables measured have different units, the relative sizes of the variances and covariances depend on the units used. Standardization overcomes this arbitrariness, giving every variable equal weight.

The next step is to find the eigenvalues and eigenvectors of \mathbf{S} . The scalar λ_k and vector $\alpha^{(k)}$ are the eigenvalue and the corresponding eigenvector of \mathbf{S} if they satisfy the equation

$$\mathbf{S}\alpha^{(k)} = \lambda_k \alpha^{(k)}, \quad k = 1, \dots, q. \quad (2.1)$$

Solving this equation for λ_k reduces to finding the q roots of a polynomial equation. If all of the roots are distinct, which is usually the case in practical problems, q different, orthogonal (since \mathbf{S} is symmetric) eigenvectors exist, corresponding to the q eigenvalues. To uniquely define the eigenvectors (apart from their sign), the orthonormality condition

$$\alpha^{(k)} \cdot \alpha^{(m)} = \delta_{km}, \quad k, m = 1, \dots, q \quad (2.2)$$

is the most convenient, with δ_{km} the Kronecker delta function.

The full set of eigenvectors $\alpha^{(k)}$ constitutes a complete orthonormal basis. As such, they can be used to expand the original data (at some time):

$$\mathbf{U} = \sum_{k=1}^q z_k \alpha^{(k)}. \quad (2.3)$$

This leads to the definition of the principal components z_k , as the projection of \mathbf{U} on the k th eigenvector:

$$z_k = \mathbf{U} \cdot \alpha^{(k)}. \quad (2.4)$$

Hence, z_k is a linear function of the original variables U_i , $i = 1, \dots, q$. In every PC, there is a coefficient for each variable, namely $\alpha_i^{(k)}$, the i th element of the corresponding eigenvector. To avoid confusion, "loadings" or "weights" is the terminology used for $\alpha_i^{(k)}$ in this paper. The acronym PC does not apply to $\alpha^{(k)}$, but to z_k . It should be noted that the sign of any PC is completely arbitrary. If every loading in z_k has its sign reversed, the variance λ_k of z_k is unchanged, and so is the orthonormality of $\alpha^{(k)}$ with all other eigenvectors.

The variances and covariances of the PCs are given by

$$\begin{aligned} \langle z_k z_m \rangle &= \langle [\mathbf{U} \cdot \alpha^{(k)}][\mathbf{U} \cdot \alpha^{(m)}] \rangle = \mathbf{S}\alpha^{(k)} \cdot \alpha^{(m)} \\ &= \lambda_k \alpha^{(k)} \cdot \alpha^{(m)} = \lambda_k \delta_{km}, \end{aligned} \quad (2.5)$$

where the orthonormality condition (2.2) has been used. It follows that z_k is uncorrelated with z_m (for $k \neq m$), and that the magnitude of the variance accounted for by z_k is λ_k . Also, the $\alpha_i^{(k)}$'s have been chosen in such a way that the variance of z_k is maximized.

In the following, we will always assume that the eigenvalues have been ordered according to $\lambda_1 > \lambda_2 > \dots > \lambda_q$. For the total variance in the dataset we can write

$$\langle \mathbf{U}^2 \rangle = \sum_{k,m=1}^q \langle z_k z_m \rangle [\alpha^{(k)} \cdot \alpha^{(m)}] = \sum_{k=1}^q \lambda_k. \quad (2.6)$$

Thus, when choosing $p \leq q$ principal components, we have $\sum_{k=1}^p \lambda_k$ of the total variance "explained."

With a selection of PCs we may approximate the time series of f_i , one of the original variables, by [using Eq. (2.3)]

$$f_i(t) = \sigma_i U_i(t) = \sigma_i \sum_{k=1}^p z_k(t) \alpha_i^{(k)}, \quad i = 1, \dots, q. \quad (2.7)$$

An attractive property is that $\lambda_k^{1/2} \alpha_i^{(k)}$ gives the correlation coefficient $\rho_i^{(k)}$ between the i th variable and the k th PC. To see this, note that $\rho_i^{(k)}$ is equal to

$$\rho_i^{(k)} = \frac{\langle z_k U_i \rangle}{\lambda_k^{1/2}}, \quad (2.8)$$

where we have used the fact that the variance of z_k is λ_k and the variance of the (standardized) U_i is one. Substitution of Eq. (2.4) yields

$$\begin{aligned} \rho_i^{(k)} &= \frac{\langle [\mathbf{U} \cdot \alpha^{(k)}] U_i \rangle}{\lambda_k^{1/2}} = \frac{[\mathbf{S}\alpha^{(k)}]_i}{\lambda_k^{1/2}} \\ &= \frac{\lambda_k \alpha_i^{(k)}}{\lambda_k^{1/2}} = \lambda_k^{1/2} \alpha_i^{(k)}. \end{aligned} \quad (2.9)$$

This property will be useful in the selection of PCs describing (parts of) the variance introduced by thermal activity. Note that $\lambda_k [\alpha_i^{(k)}]^2$ is the fraction of the variance of U_i that is explained by z_k .

b. Estimations of divergence and vorticity

The definitions of (horizontal) divergence D and (vertical) vorticity ζ are

$$\begin{aligned} D &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \\ \zeta &= \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \end{aligned} \quad (2.10)$$

with u , v the horizontal wind speed components. A simple method for estimating the local first-order derivatives from wind speed observations is to fit a "plane surface" to the data in u , x , y and in v , x , y space, respectively—the so-called planimetric method (Pedder 1981). Mathematically, these planes are represented by

$$\begin{aligned}
 u(x, y) &= a_0 + a_1x + a_2y \\
 v(x, y) &= b_0 + b_1x + b_2y \quad (2.11)
 \end{aligned}$$

where a_0, a_1, a_2 and b_0, b_1, b_2 are constants to be estimated from the data, giving us the derivative quantities as required. To solve for the constants we need at least three observations from three different points.

The solution of interest is that which minimizes the sum of the squared differences between the observations u (or v) and the values given by the equation. Using x and y values relative to the geometric center of the set of stations, the estimated coefficients of the least-squares plane may be obtained by evaluating the following expressions (m is the number of stations):

$$a_0 = \sum_{i=1}^m \frac{1}{m} u_i, \quad a_1 = \sum_{i=1}^m \beta_{1i} u_i, \quad a_2 = \sum_{i=1}^m \beta_{2i} u_i, \quad (2.12)$$

and

$$b_0 = \sum_{i=1}^m \frac{1}{m} v_i, \quad b_1 = \sum_{i=1}^m \beta_{1i} v_i, \quad b_2 = \sum_{i=1}^m \beta_{2i} v_i. \quad (2.13)$$

Here, the weights β_{1i}, β_{2i} depend only on position of the sensors: $\beta_{1i} = x_i / \sum_j x_j^2, \beta_{2i} = y_i / \sum_j y_j^2$ ($\sum_i x_i y_i = 0$ with the geometry of Fig. 2). The suffixes i and j identify a station.

For the horizontal divergence and vertical vorticity, we can now write

$$\begin{aligned}
 D &= \sum_{i=1}^m \beta_{1i} u_i + \sum_{i=1}^m \beta_{2i} v_i \\
 \zeta &= \sum_{i=1}^m \beta_{1i} v_i - \sum_{i=1}^m \beta_{2i} u_i \quad (2.14)
 \end{aligned}$$

or

$$\begin{aligned}
 D &= \boldsymbol{\beta}^{(D)} \cdot \mathbf{F} \\
 \zeta &= \boldsymbol{\beta}^{(\zeta)} \cdot \mathbf{F} \quad (2.15)
 \end{aligned}$$

The components of the vectors $\boldsymbol{\beta}^{(D)}, \boldsymbol{\beta}^{(\zeta)}$ are readily constructed from β_{1i}, β_{2i} ($i = 1, \dots, m$). The term \mathbf{F} is the stochastic vector introduced in the preceding subsection. In the case that \mathbf{F} contains parameters different from wind speed, the corresponding components in $\boldsymbol{\beta}^{(D)}$ and $\boldsymbol{\beta}^{(\zeta)}$ are zero. The inner products of Eqs. (2.15) are "linear functionals" of \mathbf{F} .

c. Estimating kinematic variables from principal components

Equations (2.15) show that divergence and vorticity can be expressed as linear functionals of \mathbf{F} . To align with the formalism of principal components, we write these variables as functionals of \mathbf{U} . To this purpose, we define a vector \mathbf{B} with elements $B_k = \beta_k \sigma_k, k = 1,$

\dots, q , such that Eqs. (2.15) become (using $U_k = f_k / \sigma_k$)

$$\begin{aligned}
 D &= \sum_{k=1}^q B_k^{(D)} U_k = \mathbf{B}^{(D)} \cdot \mathbf{U} \\
 \zeta &= \sum_{k=1}^q B_k^{(\zeta)} U_k = \mathbf{B}^{(\zeta)} \cdot \mathbf{U} \quad (2.16)
 \end{aligned}$$

After performing PCA on \mathbf{S} , we can proceed by approximating \mathbf{U} using a selection of p ($< q$) PCs, followed by the evaluation of Eqs. (2.16). However, it is easier and more straightforward to calculate D and ζ directly from the PCs, as we will show next.

We start by formulating the divergence of the eigenvectors $\boldsymbol{\alpha}^{(k)}$:

$$D^{(k)} = \sum_{i=1}^q B_i^{(D)} \alpha_i^{(k)}. \quad (2.17)$$

Combining Eqs. (2.3) and (2.16) and using the linearity of functionals yields

$$\begin{aligned}
 D &= \sum_{i=1}^q B_i^{(D)} \left[\sum_{k=1}^q z_k \alpha_i^{(k)} \right] \\
 &= \sum_{i,k=1}^q B_i^{(D)} \alpha_i^{(k)} z_k = \sum_{k=1}^q z_k D^{(k)}. \quad (2.18)
 \end{aligned}$$

With Eq. (2.17), the variance of the divergence is then given by

$$\langle D^2 \rangle = \sum_{i,j=1}^q \langle z_i z_j \rangle D^{(i)} D^{(j)} = \sum_{i=1}^q \lambda_i [D^{(i)}]^2. \quad (2.19)$$

In the same manner the corresponding expression for the vorticity can be derived.

The idea of using PCs in this manner is not only attractive for being computationally faster. The estimations of horizontal divergence and vertical vorticity as deduced from an eigenvector will be characteristic for the process in the turbulent flow that is identified by the eigenvectors.

d. Selection of principal components

The question still to be answered is which PCs are to be retained in the study. We want to distinguish signal (the important PCs) from noise (the unimportant PCs). The criterion used here, in its simplest form, is called Kaiser's rule (Kaiser 1960). It is constructed specially for use with correlation matrices and retains only those PCs whose variances are larger than one. The idea behind the rule is that if all elements of \mathbf{F} are independent, then the PCs are the same as the original variables and all have unit variance in the case of a correlation matrix. Thus, any PC with variance less than one contains less information than one of the

original variables and so is not worth retaining (note that $q^{-1} \sum_{i=1}^q \lambda_i = 1$).

Second, being interested in thermal activity we need to find the PC that identifies a significant part of the variance in the temperature signal. Therefore, the correlation coefficient with temperature was calculated by use of Eq. (2.8). This was done for every PC selected by Kaiser's rule. In the following, the PC having the highest correlation with the temperature signal has been denoted with PC_T . Note that if every one of the q PCs explains an equal part of the temperature variation, then the correlation coefficient is $q^{-1/2}$.

3. Experimental conditions and parameters

a. Site and instrumentation

The data used in this analysis were collected during the summer of 1992 at Cabauw ($51^{\circ}58'N$, $4^{\circ}55'E$) in the Netherlands. The experimental site was near the 213-m meteorological tower operated by the Royal Netherlands Meteorological Institute and has been described by Monna and van der Vliet (1987). The tower facilitates continuous measurements resulting in data files that contain half-hour averages of several parameters, including surface temperature and wind velocity and temperature at several heights.

Our measurements took place on a flat piece of grassland with the center about 225 m north of the Cabauw tower. Wind velocity measurements were made with four 3-component ultrasonic anemometers. These were mounted on top of 13-m masts positioned on the vertices of a square area with sides 80 m long ($ABDE$ in Fig. 2). Another ultrasonic anemometer-thermometer was mounted on top of a 31-m mast positioned at the center C . The different measuring height may introduce changes in variances of the velocity variable. However, using standardized variables in the PCA computations is believed to compensate for this effect. The role of the center anemometer will be discussed further in the forthcoming section.

The accuracy of the wind speed measurements is between 0.01 and 0.1 m s^{-1} , depending on the full-scale width of the measuring range. For wind direction, an accuracy of $\pm 4^{\circ}$ is assumed, while for temperature it is less than or equal to $\pm 0.1 \text{ K}$ (according to the technical specifications). Half-hour estimations (up to inversion height) of horizontal and vertical wind speed as well as temperature were made by means of a sodar and a radar.

b. Data collection and corrections

Anemometer outputs were sampled at 20 Hz. For this study we use a set of 182 half-hour runs. Measurements were performed during convective conditions. Each run was synchronized with the measurements of the Cabauw tower. Results were block aver-

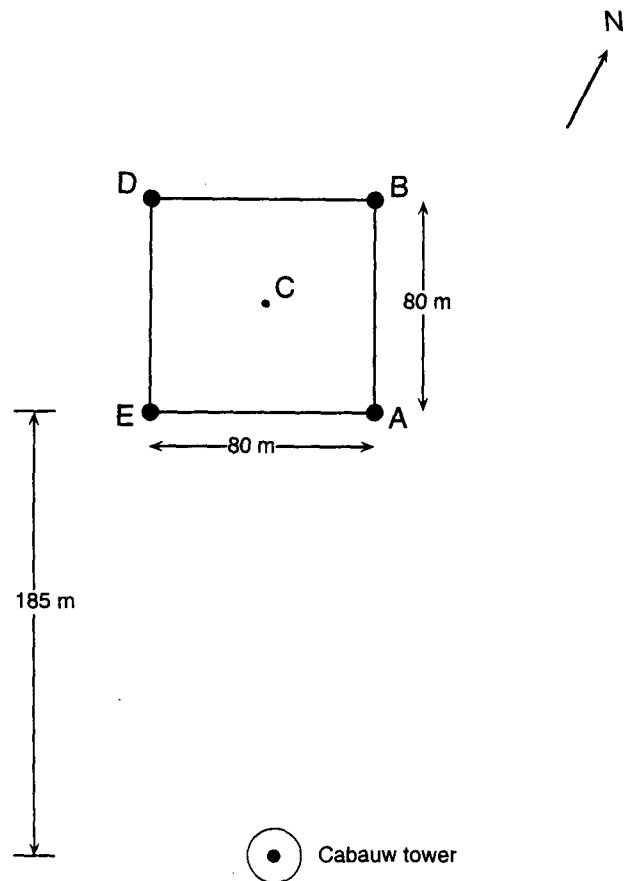


FIG. 2. Diagram of the square experimental site near the Cabauw tower.

aged over 1 s. The mean value and linear trends were removed to obtain the turbulent part of each signal.

Evaluation of divergence and vorticity obviously puts high demands on precision and accuracy with which wind velocity is measured. Therefore, several possible sources of measurement error were investigated. A possible imperfect leveling of the sonic anemometer was estimated following the method described by King and Anderson (1988). The modification of the flow by the sonic anemometer and the presence of misalignment errors were investigated according to procedures described by Wyngaard and Zhang (1985) and Weijers et al. (1994). If necessary, (small) corrections were performed on the instantaneous measurements.

c. Stability and scaling parameters

To determine the atmospheric stability, the bulk Richardson number $R_b = g(\bar{\theta}_z - \theta_0)z/\bar{\theta} \bar{U}_z^2$ has been calculated from the Cabauw tower data. Here, g is the gravitational acceleration, \bar{U}_z and $\bar{\theta}_z$ are the half-hour

averages of wind speed and potential temperature at height z , and $\bar{\theta}_0$ the half-hour-averaged surface temperature. The term $\bar{\theta}$ is the average of $\bar{\theta}_z$ and $\bar{\theta}_0$, while for z we chose 20 m.

The ratio of two velocity-scaling parameters will be used in this study: the friction velocity $u_* = [(\overline{uw})^2 + (\overline{vw})^2]^{1/4}$ deduced from the Cabauw tower data, and the convection scaling velocity $w_* = (g\bar{\theta}^{-1}H_0h)^{1/3}$. Since we only discuss the turbulent parts of variables, the prime symbol (') is generally omitted. The term H_0 is the kinematic sensible heat flux and h is the height of the first inversion. To determine h , the Cabauw tower data have been used (potential temperature profiles up to 200 m), as well as data gathered by thermal probing using sodar (up to 500 m) and radar (up to 1000 m). The behavior of u_*/w_* as function of $-R_b^{-1}$ is shown in Fig. 3. Note that $u_*/w_* = (-kL/h)^{1/3}$ (see, e.g. Stull 1988) with L the Monin-Obukhov length and k the von Kármán constant ($=0.4$). The scatter in this ratio has mainly been caused by differences in the upwind surface-layer roughness lengths (van Ulden et al. 1976). In this experiment, the higher ratios are due to larger u_* values for wind directions between 60° and 180° ($z_0 \sim 30$ cm, labeled with crosses). The lower ratios prevail when wind directions were within the (south-) western sector ($z_0 \sim 7$ cm, circles).

4. Results and discussion

a. Case study

We start with an illustrative example: a PCA applied to the data of a typical half-hour run. The variables involved are turbulent horizontal wind speed u and v (measured at masts A, B, C, D, and E) and temperature T_C (measured at location C, see Fig. 2). The positive u axis is in the direction (116°) of the half-hour-averaged wind speed (4.0 m s^{-1}). The use of the correlation matrix [**S** in Eq. (2.1)] compensates for the different variances of horizontal wind speed and temperature and for a change in variances due to the different height of measurement at location C. Each of the 11 eigenvectors has been normalized according to Eq. (2.2). To interpret the results (i.e., which flow patterns are dominant), it appears to be helpful to display the eigenvectors graphically. This has been achieved by drawing the vectorial sum of the PC weights for the horizontal wind speed components at the appropriate locations on a map. In Fig. 4 the 11 pictures can be seen. Also, eigenvalue (expressed as a percentage) and correlation with T_C [Eq. (2.9)] are given for each eigenvector as well as magnitudes of horizontal divergence and vertical vorticity [in 10^{-4} s^{-1} , estimated using Eq. (2.17)].

From Fig. 4 we deduce that the loadings for the u variables in the dominant principal component PC1 are negative at every location. Those associated with

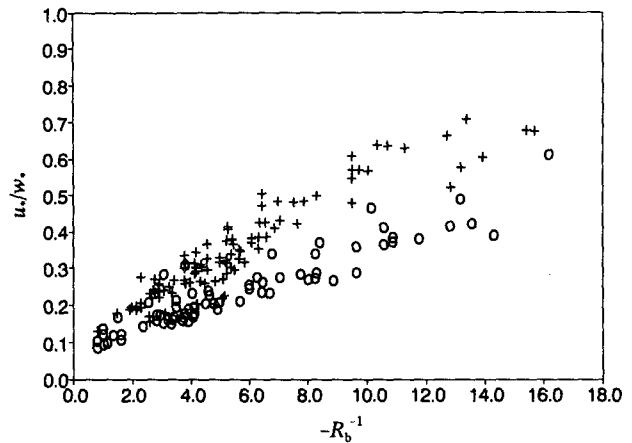


FIG. 3. Dependence of u_*/w_* on $-R_b^{-1}$. The crosses denote half-hour-averaged wind directions between 60° and 180° . The lower ratios prevail when directions are within the (south-) western sector.

the v variables are near zero at the masts positioned alongwind (A, C, and D in Fig. 2), and are of opposite sign at the crosswind stations (B and E). The loading belonging to T_C is positive. It is concluded that the dominant source of variation in this half-hour time series is between, on the one hand, periods with lower wind speeds, convergent flow fields, and higher temperatures, and, on the other hand, periods with higher wind speeds, divergent flow fields, and lower temperatures. On the whole, the variation identified is in the alongwind direction. In view of the properties of thermally generated turbulence in the SL discussed in the introduction, these periods can be associated with the passage of thermal structures. The picture reveals consistent behavior at all masts, so the events detected have scales of at least the order of the experimental square.

The next major source of variation (denoted by PC2) originates from changes in wind speed that are roughly perpendicular to the direction of the averaged background wind. In fact, its picture is largely equal to that of the first eigenvector rotated 90° clockwise, nicely demonstrating the orthogonality of eigenvectors [Eq. (2.2)]. There is considerable vorticity that will be discussed later in this section. The weight for T_C is small.

In the case of PC1 (=PC_T), PC5, PC7, and PC11, the correlation with T_C is larger than 0.30 ($q^{-1/2}$, with $q = 11$). The spatial representations of the loadings of the low-variance PC5 and PC7 are, to some extent, comparable. The increased correlation with T_C suggests that the identified variance is due to thermal activity on a smaller scale or covering parts of the square area. The weights of PC11, the principal component with the lowest eigenvalue, are (very) small except those related to the wind speed and temperature measured at mast C. It is concluded that this PC accounts for the

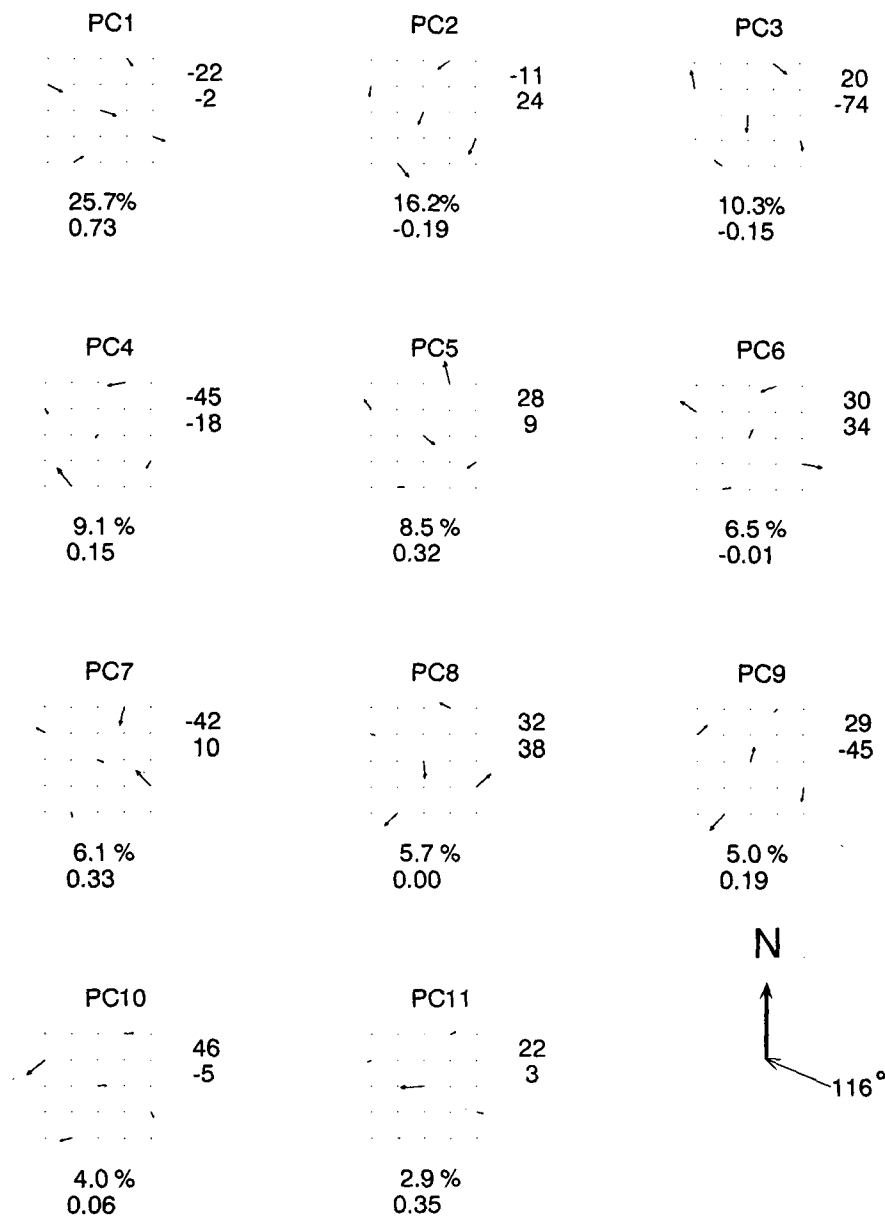


FIG. 4. Graphical representation of the 11 eigenvectors. The PC loadings belonging to the u and v components are represented by their vectorial sum drawn at the corresponding location of the five measuring stations. North is upward. The half-hour-averaged wind direction is 116° . Below each picture, variance explained by the PC and the correlation with T_c are given. At the right of each picture, estimations ($10^{-4} s^{-1}$) of horizontal divergence (above) and vertical vorticity are given.

variance introduced by small-scale thermal activity occurring at the center of the square. PCs with very small, but nonzero, variances will define near-constant relationships between a subset of variables. If one of the variables involved in such a relationship is deleted, little information is lost since its value can be determined, to some degree, from the values of the other variables. Indeed, after excluding the ve-

locity time series registered by the anemometer at mast C and repeating PCA, the dominant flow patterns hardly change. It is concluded that in the evaluation of wind fields that are of the order of the experimental square, the wind velocity at the center can be discarded. It is further noted that the different height of measurement at location C does not affect the PCA results noticeably.

The remaining PCs possess low correlations with T_C . Apparently, these identify variance in the dataset that is of nonthermal origin, that is, variance induced by mechanical turbulence.

Additional information has been obtained by doing a PCA on a dataset extended with the five time series of turbulent vertical velocity. The appearance of the two major eigenvectors remains the same. In the case of PC1, the loading for the vertical velocity at mast C has the same (positive) sign as the one for T_C . The loadings associated with vertical velocities measured at the alongwind locations are positive and large (0.32, 0.54, 0.30) compared to the coefficients at the crosswind stations (0.01, 0.07). It is concluded that the streamwise extent of the updrafts and downdrafts is much larger than their lateral extent. Indeed, Phong-Anant et al. (1981) found that the longitudinal extent of thermal structures is approximately 17 times the lateral extent. In the case of PC2, it appears that the arrows point to the crosswind station (E) for which a positive loading (0.26) has been calculated. The loading found at the oppositely located crosswind station is negative (-0.40). The three alongwind loadings show intermediate values (-0.19 , -0.22 , -0.13). It is concluded that the turbulent wind flow during these periods is always directed to the area where positive fluctuations of the vertical velocity are measured. This suggests the presence of thermal events passing nearby and explains the low correlation with T_C . Also, the positive (negative) vorticity measured when structures pass the square at the left (right) side (looking in the direction of propagation) is in agreement with earlier experimental findings (Wilczak 1984; Weijers et al. 1994).

Also informative is the reconstruction of original time series by substituting a selection of PCs in Eq. (2.7). In Fig. 5, T_C is approximated by choosing PC1, PC5, and PC7. The time trace derived from PC1 resembles the originally measured temperature behavior most (explained variance is more than 50%). Thermal structures clearly appear, but the characteristic microfronts are largely absent. Such features become apparent when PC5 and PC7 are included. This effect supports the suggestion that these PCs identify the smaller-scale thermal events. For comparison, we also give the result when only the remaining PCs are used (labeled by "environment" in Fig. 5; PC11 is not included). Each of these identify less than 4% of the variance in T_C . Therefore, in the approximation of T_C they only contribute "environmental noise."

b. General characteristics

Before presenting PCA results for the entire dataset (182 half-hour runs), we have to consider 1) the effects of a varying background wind direction on the appearance of eigenvectors, and 2) which of the PCs are retained.

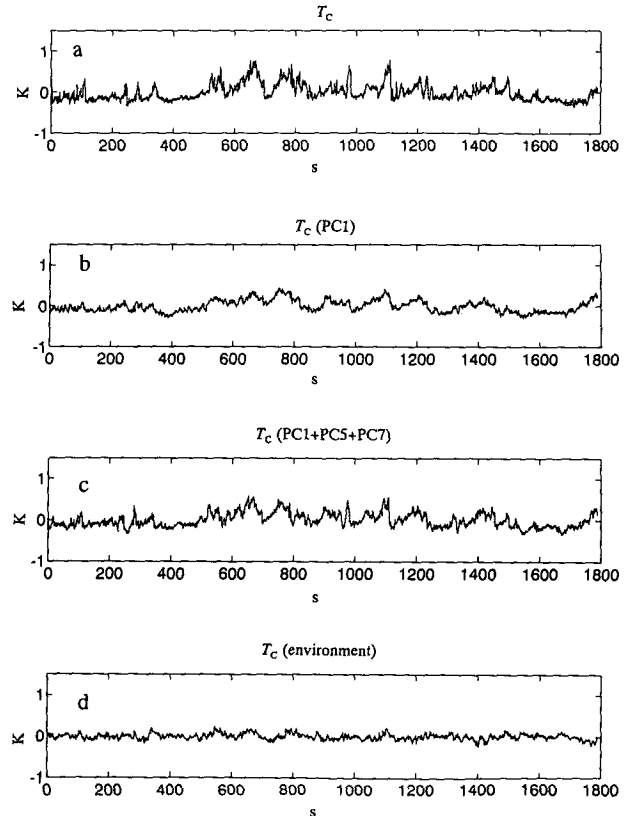


FIG. 5. Time series of temperature (K): (a) measured; (b) approximated using PC1; (c) approximated using PC1, PC5, and PC7; (d) approximated using the remaining PCs (environment). Period is 1800 s.

1) In practice, the direction of the half-hour-averaged background wind differs from one run to another. Measuring with a fixed reference system, these changes in direction affect the magnitude of the PC loadings [$\alpha_i^{(k)}$] belonging to the u , v variables. To show this, the axes of the horizontal wind speed components have been rotated over 360° with steps of 10° , using the data of the previous subsection. A PCA has been done after each step. The behavior of six PC1 loadings is shown in Fig. 6. Here, α_{u_A} denotes the weight belonging to the u variable measured at mast A. Originally, $+u$ pointed northward. After rotating counterclockwise over about 70° , all five wind speed loadings are relatively large (the minus sign is unimportant). At this angle, the u axis is practically parallel to the background wind direction (116°). Naturally, weights associated with v behave identical at angles that are 90° larger. Also note that the crosswind loadings α_{u_B} and α_{u_E} attain their extremes earlier and later, respectively, in agreement with Fig. 4. Due to the rotation, the total variance accounted for becomes split up differently across the major PCs. However, the change in variance identified by the first PC is 4% at most, while summated for the

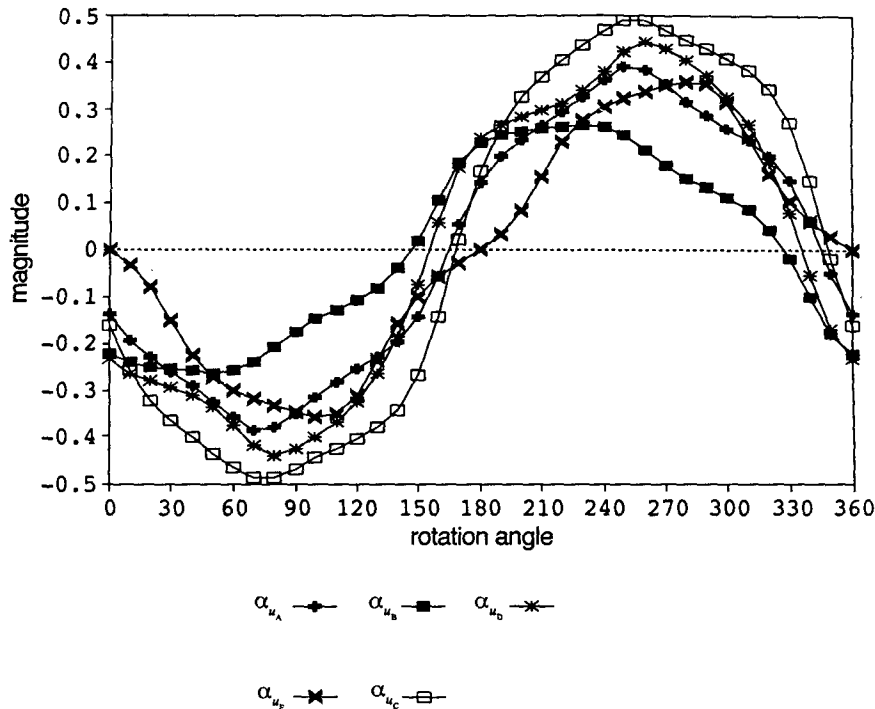


FIG. 6. The behavior of five loadings of PC1 when the reference system rotates counterclockwise with steps of 10° . The term α_{u_A} is the loading belonging to the u variable measured at mast A. At zero rotation angle, the $+u$ axis points north. When the rotation angle is 70° , the $+u$ axis points in the direction of the half-hour-averaged background wind.

three dominant PCs it is less than 1%. Concluding, in order to compare PCA findings for the various half-hour runs properly, coordinate transformations have been performed on the (perturbation) wind components yielding one component in the direction of the mean wind and another aligned with positive values to the left of the mean wind.

A second effect of the different run-averaged wind directions is that they change the orientation of the experimental square relative to the wind field. This can be understood as follows: when the wind direction is along diagonal ACD (see Fig. 2), stations B and E are positioned crosswind; when directions are 90° larger or smaller, B and E lie alongwind. Definitely, a different position in the wind field will affect the magnitudes of the PC weights. To avoid such experimental noise, runs can be selected for which, to a certain extent, the orientation of the experimental square with respect to the prevailing wind field is the same.

2) For more than 98% of the runs involved, only the eigenvalues of the first three eigenvectors are larger than one; the total variance accounted for is between 51% and 82%. (An eigenvalue of one means that roughly 10% of the variance in the dataset is explained.) Hence, we limit the discussion to the three major PCs in accordance with Kaiser's rule.

Discussing the PCA results for all the runs, we start with the variances identified by the three major PCs. These (expressed in %) are given in Fig. 7 as a function of $-R_b^{-1}$. Always, PC_T is one of the three major principal components. To discern, it has been numbered (n) according to the ordering of eigenvalues: when PC_T identifies the major source of variation, it is labeled $n = 1$ (Fig. 7a); $n = 2$ (Fig. 7b) or $n = 3$ (Fig. 7c) indicates that PC_T is the second or third dominant principal component. The crosses in Fig. 7 denote the sum of eigenvalues of the remaining two PCs. Regression lines give the linear behavior. On average, 29% of the total variance is explained by PC_T when $n = 1$, and 11% in case of $n = 3$. Simultaneously, an increase in the sum of variances identified by the two remaining PCs is measured: from 30% ($n = 1$) to 57% ($n = 3$).

Interpreting Fig. 7 in more physical terms, one observes that $n = 3$ occurs only when buoyancy is the dominant turbulence mechanism ($-R_b^{-1} < 7.0$). Also, the variance explained by PC_T increases (slightly) when the atmosphere becomes more unstable. This suggests that for $n = 3$, the PC_T s identify variation in the dataset introduced by thermal instabilities. When $n = 1$ or 2, runs have been mostly collected under less unstable conditions. On average, magnitudes of wind speed and wind shear are largest for runs with $n = 1$.

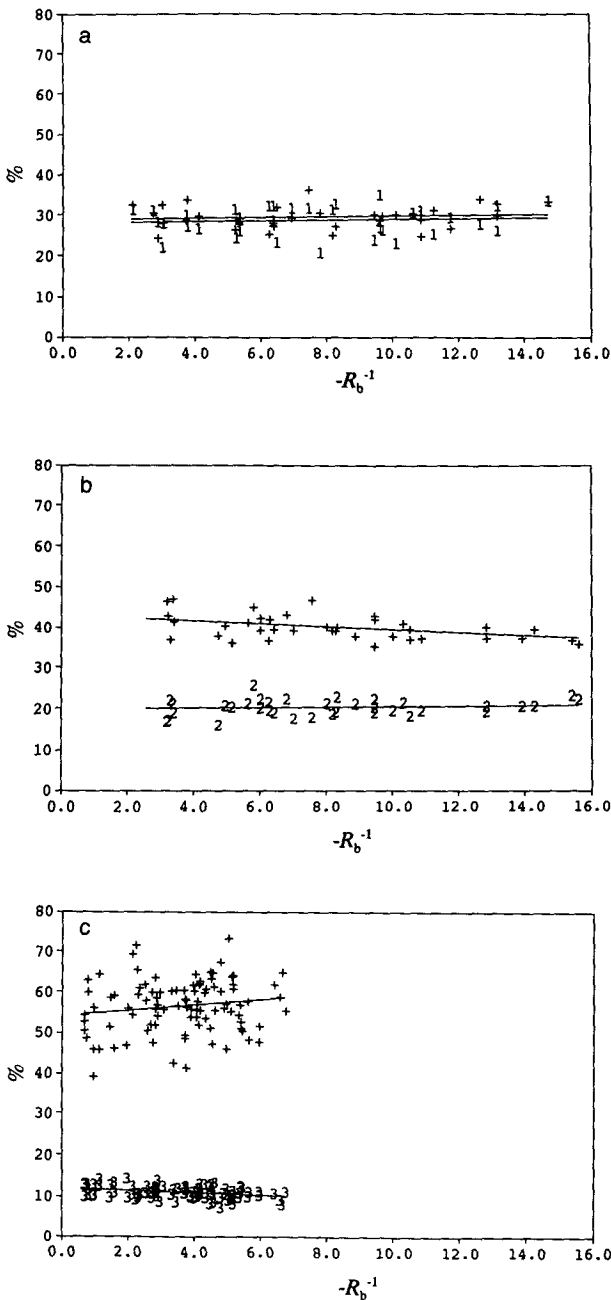


FIG. 7. The variances explained (%) by the three most important PCs as a function of $-R_b^{-1}$. When data labels are numbers, these correspond to the ranking of PC_T (see text). The crosses denote the sum of variance accounted for by the remaining two PCs.

We further concentrate on the “thermal” variance, that is, the subset of runs with $n = 3$. To find the eigenvector representations characteristic for this set, the rotated and standardized variables of each run involved have been concatenated. The standardization is necessary to give each run equal weight. Only runs for

which the half-hour-averaged background wind direction made an angle of 20° or less with one of the diagonals of the experimental square, as was motivated earlier in this section, have been used. Finally, a PCA has been performed on one data file consisting of 64 half-hour runs. The three most important eigenvectors are shown in Fig. 8 (details are as in Fig. 4). The background wind direction is from the right. It can be seen that the first and second PC identify variations in wind speed perpendicular to and parallel with the background wind, respectively. Both have low correlations with T_C . It is concluded that these PCs describe mechanically induced turbulence. The third eigenvector reveals a wind field pattern possessing (mainly lateral) convergence (divergence) when temperatures are higher (lower), typical for thermal activity (note the high correlation coefficient with T_C : 0.93).

A remarkable difference can be noted with Fig. 4. Here, the PC_T (for the subset of runs with $n = 1$) indicated the presence of a common variance in the alongwind direction of wind speed and T_C . Apparently, this PC identifies thermal turbulence influenced by the prevailing wind force. For example, the ramplike structure of plumes in the time series of T_C is a consequence of shear. In Fig. 8, however, we see that the thermal variance in the set of runs with $n = 3$ is decoupled from the (alongwind) variance in wind speed. This is likely to happen under the influence of a stronger buoyancy forcing and lower wind speeds. In this respect, also note that the picture of the first eigenvector in Fig. 4 looks like a combination of the representations of the second and third eigenvectors in Fig. 8.

5. Applications

The PC_T s have been used to estimate the kinematics associated to thermal activity in the SL. The relationship with atmospheric stability is discussed. The correlation coefficient with T_C has been assessed by use of Eq. (2.9). Its value is 0.74, on average, and the range is 0.50–0.95.

a. Kinematic properties

Horizontal divergence D and vertical vorticity ζ are estimated by first performing PCA on the 182 half-hour runs and identifying the PC_T s (accounting for a significant part of the thermal variance), and subsequently using Eqs. (2.17). Then, the absolute values of D and ζ , and their ratio have been sorted on the bulk Richardson number and block averaged over 13 successive values. This number has been chosen to obtain a standard deviation of less than half the values of divergence. Finally, results could be displayed in Fig. 9 as a function of $-R_b^{-1}$. The vertical bars denote the standard deviations.

In Fig. 9a we see that the horizontal divergence increases when instability grows ($-R_b^{-1} < 7.0$). In the

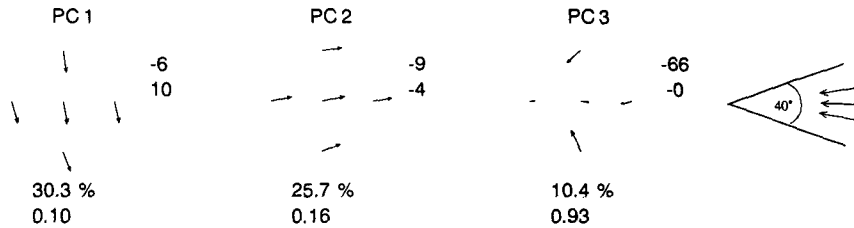


FIG. 8. The first three eigenvectors estimated for a dataset of 64 concatenated runs with $n = 3$. The directions of the run-averaged background winds were from the right ($\pm 20^\circ$). Further details are as in Fig. 4.

same range, lower values for the vertical vorticity are measured (Fig. 9b), such that with growing instability divergence becomes the dominant kinematic variable (Fig. 9c). Remarkable is the nonzero limit attained under highly unstable atmospheric conditions.

In reference to Fig. 9c, it has been investigated whether $|\zeta|/|D|$ can be parameterized. Indeed, the ratio appears to scale with a function of u_*/w_* (A. P. van Ulden 1993, personal communication), as will be shown now. The horizontal divergence can be obtained by integrating the continuity equation over z . With the incompressibility approximation, this equation reads

$$\frac{\partial w(z)}{\partial z} = -D(z). \tag{4.1}$$

Lenschow and Stephens (1980) found that near the earth's surface the convergence within thermal structures is constant with height. Using this, Eq. (4.1) reduces to

$$D = -\frac{w(z)}{z}, \tag{4.2}$$

with z the height of measurement. The absolute value of divergence can then be estimated as

$$|D| \approx a \frac{w_*}{h}, \tag{4.3}$$

where h is the height of the (first) inversion and a is a constant.

The average vorticity is of the order $u/\Delta x$ (Tennekes and Lumley 1972), where u is a characteristic velocity and Δx is a length scale (e.g., the horizontal dimension of thermal eddies). Dealing with fluctuations, we can then write for the average of the absolute value of the vertical vorticity

$$|\zeta| \approx c \frac{\sigma_{\text{hor}}}{\Delta x}, \tag{4.4}$$

with σ_{hor} the standard deviation of the horizontal turbulent velocity components and c another constant. Panofsky et al. (1977) proposed the following expression for the ratio of the standard deviation σ_{hor} to the

friction velocity u_* in the surface layer under convective conditions:

$$\frac{\sigma_{\text{hor}}}{u_*} = \left[4 + 0.6 \left(-\frac{h}{L} \right)^{2/3} \right]^{1/2}, \tag{4.5}$$

with L the Monin–Obukhov length. Using $h/L = -kw_*^3/u_*^3$ (e.g., Stull 1988) with k the von Kármán constant ($=0.4$), Eq. (4.5) can be rewritten as

$$\sigma_{\text{hor}} = (4u_*^3 + 0.33w_*^2)^{1/2}. \tag{4.6}$$

After substitution of this expression in Eq. (4.4), the absolute value of the average of the vertical vorticity becomes

$$|\zeta| \approx c \frac{\sigma_{\text{hor}}}{\Delta x} \approx \frac{c}{\Delta x} (4u_*^2 + 0.33w_*^2)^{1/2}. \tag{4.7}$$

Assuming $h \approx \Delta x$, the dependence on Δx disappears by dividing Eq. (4.7) with Eq. (4.3):

$$\frac{|\zeta|}{|D|} \approx d \left(\frac{4u_*^2}{w_*^2} + 0.33 \right)^{1/2}. \tag{4.8}$$

The ratio is now expressed in scaling variables only; d is another “constant.” If Eq. (4.8) is correct, d must remain constant when the atmospheric stability changes. Plotting d as the quotient of $|\zeta|/|D|$ and $(4u_*^2/w_*^2 + 0.33)^{1/2}$ against $-R_b^{-1}$ (Fig. 10) indeed confirms this independence. The mean value of d is $0.92 (\pm 0.52)$. As described earlier, $|\zeta|$ and $|D|$ have been estimated using $PC_{\mathcal{T}S}$. For the matter of comparison, the behavior of d is also shown with $|\zeta|$ and $|D|$ determined by using PCs that identify environmental variance. Now, d increases when $-R_b^{-1}$ decreases. In this case, $|\zeta|/|D|$ hardly changes with atmospheric stability.

Equation (4.8) also confirms the existence of the nonzero limit under highly convective conditions:

$$\lim_{u_*/w_* \rightarrow 0} \frac{|\zeta|}{|D|} \approx (0.33)^{1/2} d \approx 0.52. \tag{4.9}$$

b. Kinematic ratios in updrafts and downdrafts

Vertical velocities in thermally driven updrafts are larger (taken absolutely) than in the accompanying

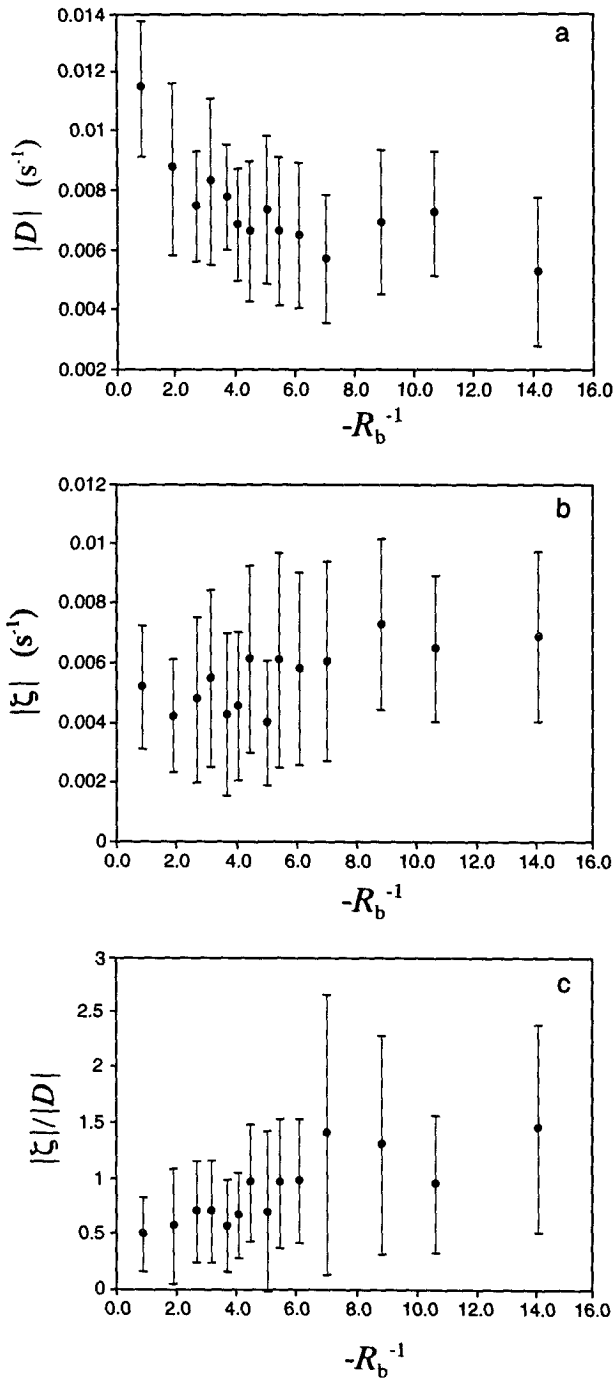


FIG. 9. Averaged magnitude (absolute value) and standard deviation (vertical bar) of (a) horizontal divergence, (b) vertical vorticity, and (c) ratio as function of $-R_b^{-1}$.

downrafts (Lenschow and Stephens 1980). Therefore, due to continuity, it can be expected that the magnitudes of divergence will be larger for updrafts than for downrafts. This is verified as follows. The time series

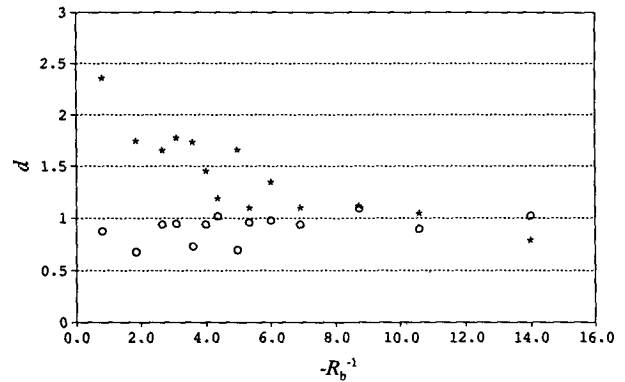


FIG. 10. Parameter d as function of $-R_b^{-1}$, estimated using PC_T (circles), and estimated from PCs that identify the environmental variance (asterisks).

of the temperature at the central mast have been approximated from Eq. (2.7) by use of the PC_T . This signal T_C^g has been used to partition the time series of divergence (also derived from PC_T) into two sets, according to $T_C^g > 0$ (assumed to identify the updrafts, further denoted by “↑”) and $T_C^g < 0$ (downrafts, denoted by “↓”). The averaged values of the ratio $|D_{\downarrow}|/|D_{\uparrow}|$

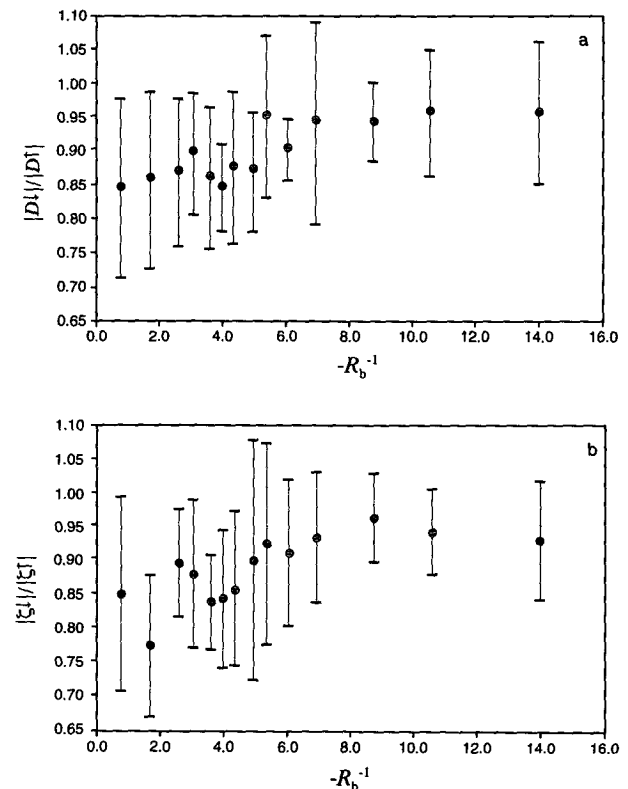


FIG. 11. (a) The ratio of horizontal divergence (absolute value) between updrafts and downrafts; (b) as in (a) but for vertical vorticity.

$|D\uparrow|$ have been calculated and displayed in Fig. 11a as a function of $-R_b^{-1}$. Indeed, the values of divergence are larger for updrafts than for downdrafts, especially when unstable atmospheric conditions prevail.

In an analog manner, vorticity can be discussed. However, the theoretical considerations are more speculative. We start with the expression for the vertical component of vorticity (Dutton 1986):

$$\frac{d\zeta}{dt} = \zeta \frac{\partial w}{\partial z} - \left(\frac{\partial w}{\partial x} \frac{\partial v}{\partial z} - \frac{\partial w}{\partial y} \frac{\partial u}{\partial z} \right). \quad (4.10)$$

The second and third term on the right-hand side are the tilting terms. We neglect the turning of the wind with height ($\partial v/\partial z \approx 0$), and we assume $\partial u/\partial z > 0$, which is usually the case. After tilting, positive vertical vorticity is generated along the $+y$ flank ($\partial w/\partial y > 0$), while negative vorticity is produced on the $-y$ flank ($\partial w/\partial y < 0$). So, the net effect is zero. The same holds for downdrafts. The first term on the right-hand side is the stretching term. Using the continuity equation, this term reduces to $-\zeta D$. Hence, the magnitude of ζ increases in time if $D < 0$. Therefore, in the case of updrafts, $|\zeta\uparrow|$ is expected to increase. Correspondingly, positive divergence (related to downdrafts) leads to vorticity reduction. The computational procedure was equal with that for divergence. In Fig. 11b, the ratios $|\zeta\downarrow|/|\zeta\uparrow|$ are shown. Again, the result supports the theoretical expectation.

6. Conclusions

Principal component analysis has been used to discover recurring intervariable associations of velocity and temperature fluctuations measured in the convective surface layer on a horizontal scale of approximately 80 m. The technique proves to be useful in distinguishing variance of thermal and nonthermal origin. For all the 182 half-hour runs investigated, there is always one principal component that identifies a significant part of the variance in the temperature time series, originating from larger-scale thermal instabilities. Coinciding with periods of higher (lower) temperatures, corresponding variance in the wind speed variables is characterized by lower (higher) wind speeds, upward (downward) movements, and convergence (divergence) in the wind fields.

Under highly unstable atmospheric conditions ($0 < -R_b^{-1} < 7.0$), the three most important principal components can be interpreted physically. The first two identify variances in wind speed in the crosswind and alongwind direction. The third principal component identifies the thermal variance. There is considerable lateral convergence (divergence) and negligible vorticity. Together, the three principal components account for about 68% of the total variance in the dataset on average.

When atmospheric conditions are less unstable ($-R_b^{-1} > 7.0$), the first principal component identifies a coupling between the thermal variance and the longitudinal wind speed variance. Apparently, the variance introduced by thermal instabilities has been modified by the higher wind speed and wind shear. The second principal component identifies variance in the crosswind direction. Smaller-scale thermal turbulence (e.g., "microfronts") are identified by PCs with lower eigenvalues.

Horizontal divergence and vertical vorticity of the velocity (variance) fields can be estimated directly from the eigenvectors when PCA is combined with a planimetric method. As expected, the magnitude of divergence, deduced from the principal component that identifies the thermal variance, is highest during the most unstable conditions. The (absolute) ratios between vertical vorticity and horizontal divergence are found to scale with a function of u_*/w_* . Both divergence and vorticity are larger for updrafts than for downdrafts, which can be explained theoretically.

Acknowledgments. We are grateful to The Royal Netherlands Meteorological Institute (J. G. van der Vliet, F. Renes, and Dr. A. P. van Ulden) for their interest, hospitality, technical assistance, and supply of the Cabauw tower data. We also appreciate comments by Drs. E. Henneken and N.-J. Bink on an earlier draft of this paper. The investigations were supported by the Working Group on Meteorology and Physical Oceanography (MFO) with financial aid from the Netherlands Organization for the Advancement of Research (NWO) under Contract 752-365-002.

REFERENCES

- Antonia, R. A., A. J. Chambers, and E. F. Bradley, 1982: Relationships between structure functions and temperature ramps in the atmospheric surface layer. *Bound.-Layer Meteor.*, **23**, 396-403.
- Dutton, J. A., 1986: *Ceaseless Wind, An Introduction to the Theory of Atmospheric Motion*. Dover Publication Inc., 617 pp.
- Godowitch, J. M., 1986: Characteristics of vertical turbulent velocities in the urban convective boundary layer. *Bound.-Layer Meteor.*, **36**, 387-407.
- Green, M. C., R. G. Floccchini, and L. O. Myrup, 1993: Use of temporal principal components analysis to determine seasonal periods. *J. Appl. Meteor.*, **32**, 986-995.
- Jolliffe, I. T., 1986: *Principal Component Analysis*. Springer Series in Statistics, Vol. XIII, Springer-Verlag, 272 pp.
- Kaimal, J. C., and J. A. Businger, 1970: Case studies of a convective plume and a dust devil. *J. Appl. Meteor.*, **9**, 612-620.
- Kaiser, H. F., 1960: The application of electronic computers to factor analysis. *Educ. Psychol. Meas.*, **20**, 141-151.
- Khalsa, S. J. S., 1980: Surface-layer intermittency investigated with conditional sampling. *Bound.-Layer Meteor.*, **19**, 135-153.
- King, J. C., and P. S. Anderson, 1988: Installation and performance of the stable instrumentation at Halley. *Br. Antarct. Surv. Bull.*, **79**, 65-77.
- Lenschow, D. H., and P. L. Stephens, 1980: The role of thermals in the convective boundary layer. *Bound.-Layer Meteor.*, **19**, 509-532.
- Monna, W. A. A., and J. G. van der Vliet, 1987: Facilities for research and weather observations on the 213 m tower at Cabauw and

- at remote locations. Scientific Report WR-nr87-5, Royal Netherlands Meteorological Institute, the Netherlands, 27 pp.
- Panofsky, H. A., H. Tennekes, D. H. Lenschow, and J. C. Wyngaard, 1977: The characteristics of turbulent velocity components in the surface layer under convective conditions. *Bound.-Layer Meteor.*, **11**, 355–361.
- Pedder, A. M., 1981: Practical analysis of dynamical and kinematic structure: Principles, practice and errors. *Dynamical Meteorology*, B. W. Atkinson, Ed., Methuen, 228 pp.
- Phong-Anant, D., A. J. Chambers, and R. A. Antonia, 1981: Spatial coherence of temperature fluctuations in the atmospheric surface layer. *Bound.-Layer Meteor.*, **21**, 465–475.
- Preisendorfer, R. W., 1988: Principal component analysis in meteorology and oceanography. *Developments in Atmospheric Science*, Vol. 17, Elsevier, 425 pp.
- Schols, J. L. J., A. E. Jansen, and J. G. Krom, 1985: Characteristics of turbulent structures in the unstable surface layer. *Bound.-Layer Meteor.*, **33**, 173–196.
- Sikora, T. D., and G. S. Young, 1993: Observations of planview flux patterns within convective structures of the marine atmospheric surface layer. *Bound.-Layer Meteor.*, **65**, 273–288.
- Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 666 pp.
- Tennekes, H., and J. L. Lumley, 1972: *A First Course in Turbulence*. MIT Press, 300 pp.
- van Ulden, A. P., J. G. van der Vliet, and J. Wieringa, 1976: Temperature and wind observations at heights from 2 to 200 m at Cabauw 1973. Scientific Report 76-7, Royal Netherlands Meteorological Institute, 30 pp.
- Weijers, E. P., A. van Delden, and H. F. Vugts, 1994: Kinematic estimates within surface-layer thermal structures. *Bound.-Layer Meteor.*, **67**, 145–160.
- Wilczak, J. M., 1984: Large-scale eddies in the unstably stratified atmospheric surface layer. Part I: Velocity and temperature structure. *J. Atmos. Sci.*, **41**, 3537–3550.
- Williams, A. G., and J. M. Hacker, 1992: The composite shape and structure of coherent eddies in the convective boundary layer. *Bound.-Layer Meteor.*, **61**, 213–245.
- Wyngaard, J. C., and S. F. Zhang, 1985: Transducer-shadow effects on turbulence spectra measured by sonic anemometers. *J. Atmos. Oceanic Technol.*, **2**, 548–558.
- Young, G. S., 1988: Convection in the atmospheric boundary layer. *Earth Sci. Rev.*, **25**, 179–198.