Processing Turbulence Data Collected on board the Helicopter Observation Platform (HOP) with the Empirical Mode Decomposition (EMD) Method

HEIDI E. HOLDER, M. ADAM BOLCH, AND RONI AVISSAR
Department of Civil and Environmental Engineering, Edmund T. Pratt School of Engineering, Duke University, Durham, North Carolina

(Manuscript received 19 October 2009, in final form 15 December 2010)

ABSTRACT

The Duke University Helicopter Observation Platform (HOP) has previously been shown to be a useful instrument for the measurement of turbulent atmospheric fluxes. As with all such measurements, especially those made from moving platforms, spurious signals, such as instrument noise and mesoscale atmospheric motions, are superposed on the desired signal. Empirical mode decomposition (EMD) is applied in a novel way to identify and separate out different signals represented by intrinsic mode functions (IMFs) in the HOP data and is shown to be an effective tool for the task. The method produces a basis that is adaptive, unique, and orthogonal, all of which are required for this type of data processing, and none of which are present in more traditional techniques. The results of applying EMD are shown to be nonlinear, and occasionally the removal of the correct number of IMFs increases the observed value of energy and fluxes calculated with the eddy correlation technique.

1. Introduction

Aircraft observations in the turbulent atmospheric boundary layer (ABL) are contaminated by aircraft movements, which are occasionally erratic and difficult to disassociate from the turbulence signals that are being measured. Additional data contamination sources include aircraft vibrations (typically producing high-frequency noise) and the nonstationarity of the atmosphere along the flight path and during the period of sampling (Lenschow 1986). Therefore, it is essential to use an algorithm to “decontaminate” the observed data series.

Traditionally, Fourier methods are used for the treatment of turbulence datasets, but they require stationarity that is often not seen in airborne data. Wavelets provide more flexibility than Fourier methods, but they require that the basis function be known a priori and that this basis function remains the same throughout the dataset. Empirical orthogonal functions (EOFs) allow for an adaptive basis, but they are not unique, so that their interpretation is not necessarily physically based. Yet for the interpretation of airborne turbulence data, it is necessary to adopt a data analysis technique that allows for nonstationarity and an adaptive basis, which is unique and orthogonal.

To observe the spatial variability of the physical and chemical properties of the atmosphere near the earth’s surface in particular, and in the entire ABL, in general, Avissar et al. (2009, hereafter A09) established a Helicopter Observation Platform (HOP). The HOP is based on a Bell 206B3 JetRanger, which is a light helicopter capable of carrying about 650 lb of scientific equipment for research missions lasting up to 3.75 h without refueling. Figure 1 shows the scientific sensors mounted permanently on the HOP nose. They measure the three-dimensional components of the wind, temperature, water vapor mixing ratio, and carbon dioxide concentration at a minimum of 40 Hz. Thus, the HOP is equipped to collect the variables needed to compute turbulent momentum, heat, moisture, and carbon dioxide fluxes (using the eddy-correlation technique) at low altitudes and low airspeeds that are not feasible with airplanes, yet are quite valuable for studying the exchanges between the earth’s surface and the atmosphere. A09 provide a thorough description of the HOP capabilities, and also use analytical, numerical, and observational studies to identify the optimal range of airspeeds at which practically undisturbed air can be
sampled in front of its nose, where the sensors and inlets are attached.

The focus of the present study is on the methodology used to process the data collected on board the HOP. It is largely based on the empirical mode decomposition (EMD) introduced by Huang et al. (1996) and further discussed by Huang et al. (1998, 1999, and 2003). The purpose of EMD is to separate a signal into its component frequencies, each of which is unique, adaptive, and orthogonal. As explained below, this method appears to be particularly appropriate for processing data collected on board aircraft. While it does not seem to have been used previously for that purpose, EMD is used together with the Fourier transform (FT). While there are advantages of using the HHT method (e.g., Huang et al. 1998; Zhaoyang et al. 2006), turbulence has been studied very extensively with FT and there is an abundance of theoretical and empirical knowledge available for comparison with our data. Finally, the pros and cons of this method are discussed.

2. The empirical mode decomposition method

Using the EMD method, any complicated dataset can be decomposed into a finite and often small number of components, which is a collection of so-called intrinsic mode functions (IMFs). An IMF represents a generally simple oscillatory mode as a counterpart to the simple harmonic function. By definition, an IMF is any function with the number of extrema and zero crossings being the same or differing by no more than one, and with its envelopes being symmetric with respect to zero (Huang et al. 1998). This decomposition method operating in the time domain is adaptive and highly efficient. Since the decomposition is based on the local characteristic time scale of the data, it can be applied to nonlinear and nonstationary processes.

The method is adaptive, which is especially important over long flights, when the data are unlikely to be stationary. Unlike the EOF expansion, which also has an adaptive basis, the results of EMD are unique and typically have a physical interpretation (Huang et al. 1998). This decomposition method operating in the time domain is adaptive and highly efficient. Since the decomposition is based on the local characteristic time scale of the data, it can be applied to nonlinear and nonstationary processes.

The method is adaptive, which is especially important over long flights, when the data are unlikely to be stationary. Unlike the EOF expansion, which also has an adaptive basis, the results of EMD are unique and typically have a physical interpretation (Huang et al. 1998).

From a practical point of view, the EMD method consists of “sifting” the data into IMFs through the application of an iterative algorithm (Huang et al. 1998; Rilling et al. 2003), as follows:

1) locate all extrema of signal $x(t)$;
2) form an envelope around $x(t)$ by interpolating between all maxima $[e_{\text{max}}(t)]$ and all minima $[e_{\text{min}}(t)]$ (typically by a spline fit);
3. The HOP sensors

For the purpose of this study, we use the data collected by two of the HOP’s “permanent” sensors, namely the Aventech Research Inc. (information online at www.ventech.com) AIMMS-20 and the Licor (information online at www.licor.com) LI-7500 (see Fig. 1). Both sensors are connected to the onboard computer—the AIMMS-20 by an RS232 serial line and the LI-7500 by two analog outputs connected to a National Instruments USB-6008 data acquisition card. The computer collects the raw data streams from both sensors and coordinates the data parsing, real-time display, and logging to files with a custom National Instruments LabVIEW (information online at www.ni.com/labview) program (A09).

The AIMMS-20 is a differential pressure-based sensor that measures the three components of the wind relative to the aircraft for various meteorological applications (e.g., Mickle 2005; Beswick et al. 2008). Other probes, including the well-known National Oceanic and Atmospheric Administration/Atmospheric Turbulence and Diffusion Division (NOAA/ATDD) Best Aircraft Turbulence (BAT) probe, are based on the same concept and theory, which are well described in the literature (e.g., Brown et al. 1983; Crawford and Dobosy 1992; Wood et al. 1997). Sensors of this type use relatively inexpensive components and are robust, a must for airborne applications (Crawford and Dobosy 1992; Beswick et al. 2008).

The AIMMS-20 consists of four modules: 1) an air-data probe located on the nose of the HOP, which senses temperature, humidity, barometric pressure, the three-dimensional aircraft-relative airflow vector and the three-axis acceleration, and magnetic field measurement; 2) an inertial measurement unit that provides three-axis acceleration and three-axis angular rates; 3) a dual-processor global positioning system that includes dual antenna inputs for differential carrier-phase measurement (one antenna is located on the nose and the other one is on the tail of the HOP); and 4) a central processing module that, among other functions, converts the inertial and GPS phase–position–velocity data into precise attitude data (roll, pitch, true heading). This processed information is shared with all other sensors on the HOP, and the AIMMS-20 is also used to coordinate the clock between the different sensors and to trigger data storage. The AIMMS-20 conducts anti-alias filtering internally by using a 50-Hz corner-frequency analog filter, then oversamples digitally at 200 Hz before a digital second-order low-pass filter is applied at 20 Hz. Finally, data are output and collected at a rate of 40 Hz.

The LI-7500 Open Path CO$_2$–H$_2$O gas analyzer consists of two components: 1) the analyzer sensor head that is mounted on the nose of the HOP and 2) the control box, which houses the electronics and is located in the aft cabin (see A09). The sensor head has a 12.5-cm open path, with single-pass optics and a large, 1-cm-diameter optical beam. Reference filters centered at 3.95 and 2.40 μm provide for attenuation corrections at nonabsorbing wavelengths. Absorption processes at wavelengths centered at 4.26 and 2.59 μm provide for measurements of carbon dioxide and water vapor, respectively. These features minimize the sensitivity to drift and dust, which can accumulate during normal operation. The LI-7500 is operated without the use of LICOR software, which typically gives output rate options of 5, 10, or 20 Hz, and instead, the raw data are collected through the analog output of the Licor control box at an output rate of 160 Hz, which is reduced to 40 Hz for temporal coordination with the AIMMS-20 data. This reduction, combined with the removal of the highest-frequency mode, IMF$_0$ (as discussed in section 6), also serves as a manual anti-alias filter for the LI-7500 data.

4. Data preprocessing

It can be easily understood from differential pressure theory as described in the literature mentioned in section 3 (e.g., Brown et al. 1983; Crawford and Dobosy 1992; Wood et al. 1997) that the position of the AIMMS-20 in

$$m(t) = [e_{\max}(t) + e_{\min}(t)]/2;$$

(1)

4) subtract the envelope mean from the signal to yield the first component, $h(t)$, sometimes also called the detail, as

$$h(t) = x(t) - m(t);$$

(2)

5) repeat steps 1–4 iteratively with the detail $h(t)$ replacing the signal $x(t)$ until the resulting $h(t)$ meets the aforementioned criteria that define an IMF as

$$h_{n+1}(t) = h_n(t) - m_n(t);$$

(3)

and

6) sift out the IMF $h(t)$ and restart the process with the remaining data [which is the mean $m(t)$ from steps 3 and 4 in the last iteration].

This algorithm generates a finite number of IMFs, with the number of extrema decreasing with each mode (Rilling et al. 2003). It should be noted that using the usual HHT notation, the first result of EMD contains the highest-frequency oscillations, and is referred to as IMF$_0$ (Lundquist 2003). In the notation used here, the IMF containing the lowest-frequency mode is IMF$_k$, the next lowest-frequency mode is IMF$_{k-1}$, etc.
flight relative to the local Cartesian coordinate system is crucial for obtaining accurate readings of the three components of the wind. Thus, a series of calibration flights were conducted to correct for the effects of the HOP attitude when flying at different airspeeds and in crosswinds. The calibration procedure (Beswick et al. 2008) includes a flight in unchanging, low wind speed where 1) the yaw is purposefully set 10° to port, then to 0°, then 10° to starboard while flying at a range of forward airspeeds; and 2) similar alterations in pitch and/or a rapid climb and descent are performed. The manufacturer can then analyze the calibration flight data (e.g., steps 1 and 2 above) to evaluate the sensor’s response to sideslip and angle of attack, respectively and provide appropriate calibration coefficients to be used in the first step of data preprocessing. It should be noted that while, theoretically, there should be no need to repeat the calibration flights as long as the position of the AIMMS-20 on the HOP remains unchanged, these calibration flights are performed before major field campaigns to ensure the accuracy of the algorithm and to adjust for any minor alteration in the instrument attitude that may have occurred.

The LI-7500 data are calibrated for water vapor and carbon dioxide concentration according to a procedure described by its manufacturer (LI-7500 manual available online at www.licor.com). For the calibration procedure, the LI-7500 is connected via the serial port to the computer running the Licor calibration software, and a calibration tube is connected to the sensor head, across the sampling area. Zero and span gases for carbon dioxide and water are sent through the calibration tube, and the Licor software provides the calibration coefficients. This procedure is repeated before each field campaign to account for any instrument drift over time. As part of the aforementioned AIMMS-20 calibration flights, the LI-7500 was oriented in different directions relative to its mount (Fig. 1), including parallel and perpendicular to the AIMMS-20 (and, therefore, the flight direction). However, no noticeable impacts of orientation were identified and there is no specific preprocessing procedure required for that sensor.

Occasionally, during a high wind gust that may cause the HOP to pitch, roll, and/or yaw excessively and rapidly, the attitude of the HOP is outside the range in which the calibration algorithm is applicable. Indeed, due to the aerodynamics of the hemispherical differential pressure sensor, the optimal operating range is within a sideslip of ±5°. Therefore, as part of our data preprocessing, an algorithm that automatically flags the data for such occurrences, which are typically not more than a few seconds long, was developed. To maintain the continuity of the data series, this algorithm substitutes the anomalous data with randomly generated data that share characteristics of the first and second statistical moments with the immediately previous and subsequent 0.1–0.25 s of data. Specifically, the replacement data have a variance equal to the average variance of the surrounding data, and its mean follows a linear interpolation between the means of the surrounding data. While not perfect, it seems that the error incurred by this substitution is much less significant than the error present in the flawed data, especially when such events are only occasional, as shown in the example in the next paragraph. For similar reasons, the data collected during steep turns, rapid climbs, or descents are typically ignored.

A very large disturbance is incurred in the temperature data collected with the AIMMS-20 when the HOP radio is activated for communication. Similarly, when flying near radio and TV towers, data are contaminated. These disturbances are easily identified in the time series, and the dataset is correspondingly adjusted as described in the previous paragraph. These temperature anomalies typically show a drop on the order of about 25 K. Statistics from a typical temperature disturbance and the surrounding undisturbed data demonstrate the possible effects of such an anomaly, and justify our correction technique. In the example data (chosen from an earlier flight on the same day described in the next section), the temperature disturbance is 1.7 s long, and is compared to the immediately previous and subsequent 10 s of undisturbed data. Before the disturbance, the data mean is 295.02 K and its variance is 0.021 K², and after the disturbance, the mean is 295.12 K with a variance of 0.011 K², while the anomaly itself has a mean of 270.91 K and a variance of 1.52 K². Taking into account the full example section of data (1.7 s outlier along with 10 s before and after), the raw data have a mean of 295.20 K, which is 2 degrees too low, and an unrealistic variance of 42.64 K². After applying the correction detailed in the previous paragraph to the anomalous data, the example data section has a mean of 295.11 K and a variance of 0.014 K², which gives much better consistency with the surrounding, undamaged data. Although some flux information is undoubtedly lost by this method, the length of time affected is very short (in this example, about 0.3% of the total flight leg time), and the correction benefits greatly outweigh the errors introduced.

The AIMMS-20 sensor protrudes about 0.25 m ahead of the LI-7500 (Fig. 1). A09 explains that for research missions aimed at measuring the atmospheric variables needed to calculate fluxes using the eddy-correlation technique, it is most practical to fly the HOP at a 30 m s⁻¹ airspeed. This speed is slow enough to sample the atmosphere at a high resolution yet fast enough to perform a reasonably long flight leg. It is also an optimal speed for the HOP sensors to remain clear of the main-rotor wake...
while simultaneously avoiding the distortion of airflow due to its airframe (see analytical, numerical, and observational studies in A09). At 30 m s\(^{-1}\), the distance between the AIMMS-20 and the LI-7500 is flown in ~8 m s. One could suspect that this discrepancy could have an impact on the covariances and cospectra between the wind components and the water or carbon dioxide concentration, so studies were conducted to assess any alteration in correlation between water concentration data (as output by the LI-7500 and as derived from AIMMS-20 measurements) by slightly shifting in time one instrument’s data relative to the other. However, no consistent effect could be detected by these time shifts, and it was determined that any effects of the position difference cannot be captured by the current system. This is not unexpected, since the instrument response time for the AIMMS-20 is 25 ms, and the typical instrument response time for the LI-7500 is 50 ms. The electronic reaction time between the actual sampling and data registering could be a factor that similarly affects such covariances and cospectra. Due to the inconclusive results of the aforementioned time-shift studies, estimating any potential bias is problematic. Because of this, the current system relies on the manufacturers’ information regarding the instrument delays.

When sampling an atmospheric variable that is stratified vertically, altitude variations introduce synthetic disturbances into the relevant data series. Pressure is the variable most obviously always stratified but the other variables may also depict strong vertical gradients. The EMD method is used to correct the data collected for this type of undesired, artificial perturbation, as shown below.

5. Flight description

A flight performed during the Cloud and Land Surface Interaction Campaign (CLASIC) on 19 June 2007 is used to illustrate the issues involved in collecting turbulence data on board the HOP and to demonstrate the use of the EMD method. As illustrated in Fig. 2, a triangular pattern, 10 km on each side, was flown at different altitudes. In addition, two “profiling” flight segments aimed at characterizing the vertical structure of the atmosphere were performed. The vertical climb rate during such profiling segments is about 100 m min\(^{-1}\) and, therefore, the data collected are not typically used for flux calculation. However, the real-time display of potential temperature is accurate enough to provide a good estimate of the top of the convective boundary layer (CBL), which is then used to establish the various altitudes at which the triangular patterns are flown.

For field campaigns such as CLASIC, geographic locations for each day and a general flight plan (i.e., that the flight will include several triangles at different altitudes with a few vertical profiles interspersed) are usually determined before takeoff. However, the specific details of the flight pattern, particularly how many triangles are flown and at which altitudes, are commonly chosen during the flight and depend strongly on the CBL height determined by the real-time potential temperature display mentioned above. Typically, if the CBL is less than 300 m high, as could be the case in early morning flights, three triangles are performed: one near the ground surface, one near (i.e., 30–50 m below) the top of the CBL, and one near the middle of the CBL. When the CBL is deeper, the near-surface and near-top CBL triangles are still flown, with more triangles flown in between these levels with a separation of 100–150 m.

Figure 3 depicts the time series of the flight altitude (above sea level) and of the preprocessed wind components and scalars collected by the AIMMS-20 and the LI-7500 during a flight made 19 June 2007, during CLASIC. The altitude provides a clear history of this specific flight. During the first ~800 s, the HOP climbed by ~100-m-high “steps,” moving from near the ground surface up to just below the top of the CBL, and maintaining a nearly constant altitude for about 2 min for each of the steps. While longer legs help reduce the magnitude of the error (see A09), a 2-min leg seems to provide sufficient data to estimate the turbulence fluxes reasonably well. At the end of this pattern, the HOP climbed through the top of the CBL (which is identified by the real-time display of the potential temperature profile in the helicopter cockpit), up to above the entrainment layer to provide information on the location of the inversion layer capping the CBL. Next, three triangles were flown as explained above (i.e., one near the top, one near the middle, and one near the bottom of the CBL). Data from the first leg of the lowest-altitude triangle are shown in Fig. 4, using the
same conventions as in Fig. 3. Given that the HOP is flown at a more or less constant airspeed of 30 m s\(^{-1}\) (turbulence causes some small airspeed fluctuation) through a 30-km triangular pattern, each triangle should be completed in 1000 s. However, wind speed variations can lead to differences in the amount of time required to complete a triangle, as evidenced by the fact that in this particular case, the three triangles were flown in about 47 min, as opposed to the predicted 50 min. Following these three triangles, a continuous climb and descent pattern was flown to profile the CBL and to assess possible changes in CBL height. Figure 5 shows the top of this profile as a plot of virtual potential temperature against altitude, and indicates an inversion capping the CBL at approximately 720 m MSL (\(\sim 400 \text{ m AGL}\)). Then, two additional triangles were flown at intermediate altitudes to complete the full characterization of the CBL.

The collected data clearly illustrate the greater intensity of turbulence near the ground surface, which is generated by wind shear and buoyancy. This is particularly evident in the time series of the scalars. Furthermore, there is a visible correlation between altitude and the scalars, showing a decrease in temperature with height as expected in a CBL. Moisture and carbon dioxide also generally decrease with height although the carbon dioxide concentration very close to the surface increases slightly with height. This is consistent with the negative mean carbon dioxide flux at the lowest flight level, which indicates absorption of carbon dioxide by the land surface. The large moisture source at the surface (i.e., evapotranspiration) is consistent with the relatively low height of the CBL (\(\sim 400 \text{ m}\); see Fig. 5), which is a consequence of a relatively low contribution of buoyancy (sensible heat flux) and strong latent heat flux. June 2007 was the wettest month of June on record in Oklahoma and, therefore, these observations are not surprising. The entire flight displayed in Fig. 3 was performed between 1345 and 1500 LT. At this time of the year, when the winter wheat crops are senescent and the summer corn crops are not yet growing, the vegetation photosynthesis activity is low and, correspondingly, the sink of carbon dioxide is small.

Figure 6 presents the spectra of atmospheric variables obtained in the first leg of the third triangle (see also the raw data in Fig. 4). The reason for selecting this specific leg is that the near-surface flights are the most challenging to process, given the variations of altitude due to topographical features as well as various obstacles on the ground (e.g., power lines, trees, houses, etc.). Furthermore, wind shear near the ground affects the stability of the HOP. Even more importantly, observations in the near-surface layer are the special niche of the HOP among airborne platforms (see A09).

A few characteristics relevant to the HOP and its current set of sensors are worth mentioning. For instance, one can notice that due to sensor limitations, the highest frequency of valuable data that can be used for our studies is \(\sim 10 \text{ Hz}\). The main rotor of the JetRanger, which has two blades, has a constant 396 rpm \([\pm (1-2)\%]\) that generates the disturbance peak seen in all spectra at \(\sim 13 \text{ Hz}\). All spectra show an inertial subrange with a slope of \(-\frac{5}{3}\) up to sensor limit at \(\sim 10 \text{ Hz}\), as expected from Kolmogorov theory. The one exception is the wind spectra, which shows a slope less than the expected \(-\frac{5}{3}\). This phenomenon is seen consistently in only the lowest-altitude flight legs, when the HOP is operating just above the ground level or canopy top. In flight legs collected
anywhere in the ABL other than the lowest altitude, the $w$ wind spectra follow a $-5/3$ slope quite consistently. This could be the result of a relatively greater contribution to the turbulent vertical winds in the higher-frequency range, possibly due to the close proximity to the ground surface and more influence from turbulence intensity from smaller-scale shear than larger-scale buoyancy in this altitude range, which may lead to the appearance of a “flatter” spectral slope. It is also possible that the spectral slope is influenced by the fact that in the lowest-altitude flight legs, the HOP flies close to the ground, following the terrain when possible, and increasing altitude for power lines, etc., which could explain the differences from spectra collected from low-altitude towers (e.g., Kaimal et al. 1972). Further study is currently under way to explore this issue, and a more complete analysis using data from multiple flights during the CLASIC field campaign will be presented in another paper (M. A. Bolch et al., unpublished manuscript).

A sensor with a response time at higher frequencies than the AIMMS-20 would possibly show a continuation of the $-5/3$ slope at higher frequencies. But as explained in A09, flying near the ground surface is particularly destructive for the sensors (due to collisions with insects and dust) and the choice of a robust sensor comes with the loss of higher sensitivity. As explained below, given the negligible impact that higher frequencies have on the calculation of the turbulence fluxes, this is an acceptable compromise for these types of missions.

Some of the cospectra between the wind components and the temperature, the water mixing ratio, and the carbon dioxide concentration are presented in Fig. 7. We find that the subrange slopes of the cospectra are close to $-5/3$. While Lumley (1964), Kaimal et al. (1972), and Kader and Yaglom (1991) obtained subrange slopes of $-7/3$, Wyngaard and Coté (1972) report a $-3$ slope, and, similarly to this study, Van Atta and Wyngaard (1975), Wyngaard et al. (1978), and Antonia and Zhu (1994) observed a $-5/3$ slope. A more detailed analysis of spectra and cospectra obtained during the CLASIC field campaign including multiple flights and different experimental sites is currently under way, and will be published.
in a later manuscript (M. A. Bolch et al., unpublished manuscript).

6. Data processing

Figures 8–10 show some of the IMFs obtained for the vertical component of the wind, the temperature, and the water vapor concentration, respectively. Specifically, the figures show the highest-frequency IMF, IMF_0, and the four lowest-frequency IMFs, IMF_{-1}, IMF_{-2}, IMF_{-1}, and IMF_{-1}, as calculated from the first leg of the lowest-altitude triangle. This analysis was also performed for the other variables and for all flight legs, but given that they do not contribute additional insights relevant for this study, for brevity, they are not presented here.

Here, IMF_0 represents the high-frequency noise that is associated with the sensor and has no physical meaning. The spectra of IMF_0 show flat (zero) slopes, which indicates that the information contained in this highest-frequency IMF is predominantly noise. Often, this can be seen in the spectra of the raw data as well; for example, note the near-zero slope region on the high-frequency end (about 10 Hz and higher) of the temperature spectra in Fig. 6. The flattened slope of the IMF_0 spectra is not always evident in the raw data spectra, though, because due to the adaptive nature of EMD, each mode contains a range of frequencies, so the contributions from IMF_1 may have an influence in the IMF_0 range of the raw data spectra. Another reason to remove IMF_0 is that this mode also contains the signal of the main rotor of the JetRanger at 13 Hz, which can be seen as a spike in the same figure, particularly in the spectra of the winds. The elimination of IMF_0 from the original time series is straightforward and has no practical impact on the variables, their variances, and their covariances (i.e., kinematic fluxes). In fact, as can be seen in Table 2 (described in more detail later), which summarizes the impacts of the removal of IMFs on turbulent flux calculations, removing IMF_0 changes the sensible heat and carbon dioxide fluxes by only about 1%, the latent heat flux by only about 0.1%, and the TKE by 3%.

The residual of the EMD algorithm, which is the lowest-frequency mode (IMF_{-1}), contains the general trend of the time series and its elimination has the effect of detrending the collected data. Its impact on the variances and covariances is often quite significant. In addition, it is often apparent that this IMF must be removed, along with IMF_{-1}, because they complete only about two full cycles, or less, throughout each flight leg. This is not adequate to resolve covariances from a statistical standpoint, so their inclusion is meaningless. Physical justification is considered necessary for the elimination of any other IMFs, and since it is not always straightforward to relate an IMF to

![Fig. 5. Virtual potential temperature (K) vs altitude (m, MSL) from data collected on HOP in a flight performed during CLASIC on 19 Jun 2007. Black dots indicate the raw data, and the gray line shows a 20-s moving average. The inversion indicates the top of the CBL at approximately 720 m above sea level (MSL), which corresponds to about 400 m AGL.](image-url)
the physics of the observed variable, this is the most obvious weakness of the method. Similarly to any other data processing technique, some subjectivity is used in the process.

A correlation analysis is used to help in the decision to remove specific IMFs. As emphasized above, after detrending the time series by removing IMF$_i$, it is anticipated that eliminating fluctuations caused by (i) non-stationarity of the atmosphere observed during the flight leg (possibly in addition to the series' lowest-frequency component, depending on the distance-flight time of the leg), (ii) altitude variations, and (iii) pitch, roll, and yaw (attitude) movements of the HOP will be the most important corrections required to be made to the dataset. The altitude and attitude changes of the HOP are inspected in this correlation analysis because such movements through areas of the ABL with strong stratification could lead to spurious contributions to flux calculations. Since such fluctuations are expected to occur over time scales of at least a few seconds, the low-frequency IMFs are expected to carry these fluctuations.

Table 1 presents the correlation analysis between these parameters and the progressive sum of IMFs (i.e., IMF$_i$, IMF$_i$ + IMF$_{i-1}$, IMF$_i$ + IMF$_{i-1}$ + IMF$_{i-2}$, etc.) for the time series of some of the atmospheric variables obtained during the first leg of the third triangle of our flight. For example, the first line of the table shows the $r^2$ (square of the Pearson’s correlation coefficient) value as calculated from the altitude of the platform and the successive IMFs of $u$. The minimum $r^2$ value necessary for a correlation to be significant at the 95% confidence level, for a dataset of this size, is 0.04. At the 99% confidence level, the minimum $r^2$ is 0.09. Based on this straightforward analysis, an $r^2 > 0.10$ is taken to be justification for the elimination of a particular IMF for this flight leg. It is important to note that this value is chosen on a leg-by-leg basis, and the exact value is determined not only by the significant correlation level but also by careful examination of the data and the IMFs. In this case, eliminating IMF$_{i-1}$ (in addition to IMF$_i$ and IMF$_0$ as explained earlier) is justified for $u$, and further eliminating IMF$_{i-2}$ is justified for $v$, $w$, and $T$. The last four IMFs of temperature exhibit a correlation with yaw, so these IMFs should be eliminated as well. This correlation is likely due to the aerodynamics of the sensor mount, which is designed to protect the fragile thermistor.

Figure 11 illustrates the impacts of removing various IMFs from the data on the variable spectra. Although Fourier spectra are not an ideal tool for the evaluation of turbulent data, their use is well prescribed, and, therefore, examination of the effects on the spectra of removing IMFs may provide additional insights. Colored lines represent the removal of subsequent IMFs on both the high- and low-frequency ends of the spectra. The removal of IMFs uniformly has the expected effect on the spectra,
reducing the spectra in that frequency range more with each IMF removed. Table 2 summarizes how the removal of IMFs affects turbulent fluxes in the CBL. The first row shows the fluxes as calculated from the raw data, and the second row shows the resulting fluxes when IMF_0 is removed from both variables. The third row in Table 2 shows fluxes calculated from data with one low-frequency IMF removed from each variable, and subsequent rows have additional IMFs removed from both variables before the flux calculation. It is interesting to note that the impacts of the removal of subsequent IMFs is quite non-linear and that it is not possible, a priori, to estimate the impacts of eliminating a particular IMF on the fluxes. Indeed, in some cases the removal of a specific IMF decreases the variance of a given variable while the following IMF might increase it. From the results presented in Table 2, the removal of successive IMFs decreases the magnitude of both the carbon dioxide flux and the TKE, as might be expected. However, the removal of the last two IMFs increases the magnitude of the sensible heat and latent energy fluxes. This result is not entirely surprising, as these IMFs are correlated with altitude and aircraft attitude. These erroneous signals are superposed on the physical turbulence signals, so their removal improves the quality of the data. This in turn improves the correlation between variables.

The “best estimate” flux values in Table 2 are calculated by filtering low-frequency IMFs for each variable according to the correlation analysis as summarized previously in this section and using the $r^2$ values shown in Table 1. For example, according to Table 1, three low-frequency IMFs should be filtered for $w$ due to correlations with altitude and pitch, and four low-frequency IMFs should be filtered for $T$ due to correlations with yaw. Therefore, the best-estimate calculation of the sensible heat flux uses $w$ with IMF_3, IMF_{-1}, and IMF_{-2} removed as well as $T$ with IMF_3, IMF_{-1}, IMF_{-2}, and IMF_{-3} removed. This explains why the best-estimate flux value for sensible heat

![Figure 7](image)

**FIG. 7.** As in Fig. 6, but cospectra between wind components, temperature, water vapor concentration, and carbon dioxide concentration.

![Figure 8](image)

**FIG. 8.** Graphical representation of (top to bottom) IMF_0, IMF_{-3}, IMF_{-2}, IMF_{-1}, and IMF_{-3} obtained for the vertical component of the wind collected on HOP in one of the flight legs performed during CLASIC on 19 Jun 2007.
falls between the values calculated by removing three IMFs from both variables and that calculated by removing four IMFs from both variables. Three IMFs are the optimal number removed for $w$ and two is ideal for carbon dioxide, yet the combination of these filters in the best-estimate flux calculation yields a flux value between that of three and four IMFs removed from each variable, further illustrating the nonlinearity of the technique. For the latent heat flux, the optimal number of low-frequency IMFs to remove for moisture is two, and for $w$ the optimal filtering is at three low-frequency IMFs, which results in an unsurprising best-estimate flux value between those calculated by removing two and three IMFs from both variables. Overall, the calculations reveal the subtle nature of EMD filtering, and that much care should be taken to analyze particular circumstances for any measurement flight.

Another type of fluctuation that can be seen in our dataset is that triggered by vibrations of the HOP. As explained above, the effects of the main rotor (which includes various vibrations) is well visible in all spectra. Other rotating components of the helicopter (e.g., the

---

**Table 1. Correlations ($r^2$) between IMFs of atmospheric data collected on the HOP on 19 Jun 2007 and the HOP altitude and attitude (pitch, roll, and yaw). Highlighted in boldface are all $r^2 > 0.10$.**

<table>
<thead>
<tr>
<th></th>
<th>IMF$_0$</th>
<th>IMF$_1$</th>
<th>IMF$_2$</th>
<th>...</th>
<th>IMF$_{-4}$</th>
<th>IMF$_{-3}$</th>
<th>IMF$_{-2}$</th>
<th>IMF$_{-1}$</th>
<th>IMF$_{r}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$ and $z$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.01</td>
<td>0.06</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td>$v$ and $z$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.01</td>
<td>0.00</td>
<td><strong>0.28</strong></td>
<td><strong>0.30</strong></td>
<td><strong>0.03</strong></td>
</tr>
<tr>
<td>$w$ and $z$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>0.19</strong></td>
<td><strong>0.08</strong></td>
<td><strong>0.19</strong></td>
</tr>
<tr>
<td>$T$ and $z$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.01</td>
<td>0.00</td>
<td><strong>0.30</strong></td>
<td><strong>0.40</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>$q$ and $z$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.01</td>
<td>0.06</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>$C$ and $z$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.05</td>
<td>0.03</td>
<td>0.00</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>$u$ and roll</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.04</td>
<td>0.08</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>$v$ and roll</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>$w$ and roll</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>$T$ and roll</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>$q$ and roll</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>$C$ and roll</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.12</td>
<td>0.03</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>$u$ and pitch</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.04</td>
<td>0.01</td>
<td>0.03</td>
<td>0.10</td>
<td>0.02</td>
</tr>
<tr>
<td>$v$ and pitch</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.07</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>$w$ and pitch</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>0.13</strong></td>
<td><strong>0.18</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>$T$ and pitch</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>0.13</strong></td>
<td><strong>0.18</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>$q$ and pitch</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td><strong>0.10</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>$C$ and pitch</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$u$ and yaw</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.01</td>
<td><strong>0.21</strong></td>
<td><strong>0.03</strong></td>
<td><strong>0.21</strong></td>
</tr>
<tr>
<td>$v$ and yaw</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.05</td>
<td><strong>0.40</strong></td>
<td><strong>0.01</strong></td>
<td><strong>0.40</strong></td>
</tr>
<tr>
<td>$w$ and yaw</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td><strong>0.20</strong></td>
<td><strong>0.31</strong></td>
</tr>
<tr>
<td>$T$ and yaw</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td><strong>0.24</strong></td>
<td><strong>0.13</strong></td>
<td><strong>0.28</strong></td>
<td><strong>0.00</strong></td>
</tr>
<tr>
<td>$q$ and yaw</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td><strong>0.11</strong></td>
<td><strong>0.18</strong></td>
</tr>
<tr>
<td>$C$ and yaw</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>...</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td><strong>0.14</strong></td>
<td><strong>0.00</strong></td>
</tr>
</tbody>
</table>
engine, the transmission, and the tail rotor) have RPMs that are much faster than the rotor and, therefore, are likely to have vibration peaks at much higher frequencies than those detected by our sensors. Here, it appears that all these vibrations are eliminated from the time series by subtracting the respective IMF0 from them.

7. Conclusions

A typical (yet complex) research flight conducted with the HOP during the CLASIC experiment demonstrates the potential of the EMD method for airborne data processing. The capability to detrend the time series for nonstationarity and to eliminate the contamination that is typically found in airborne data, especially when flying in a stratified atmospheric layer is demonstrated. The main advantage of this method is that it is empirically based and does not need to make any particular assumptions about the data structure. In addition, EMD is capable of handling nonlinear, nonstationary data, which makes it ideal for the analysis of atmospheric turbulence data. It is an especially useful technique for the analysis of HOP data, as it can identify and separate the signals due to vibrations and platform motions that have changing frequencies and amplitudes. Traditional Fourier methods use only a single frequency or a range of frequencies at constant amplitude and the frequency chosen must be a harmonic of the sampling frequency. In addition, Fourier methods are inappropriate for nonstationary data. Wavelet methods are also limited to a single frequency range and amplitude, both of which must be known a priori. Unlike the empirical orthogonal function (EOF) expansion, which also has an adaptive basis, the results of EMD are unique, so that each one has a physical interpretation (Huang et al. 1998).

Analyzing the impacts of the low-frequency IMFs on the correlations between the atmospheric variables being measured and the aircraft’s altitude, pitch, yaw, and roll helps eliminate disturbances in the time series that are

### Table 2. The impacts of EMD filtering on surface turbulent fluxes measured by HOP on 19 Jun 2007. Best-estimate values are calculated using Table 1 values.

<table>
<thead>
<tr>
<th>IMFs removed</th>
<th>SH (W m⁻²)</th>
<th>LH (W m⁻²)</th>
<th>CF (mg m⁻² s⁻¹)</th>
<th>TKE (m² s⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (raw data)</td>
<td>60.1</td>
<td>118.5</td>
<td>-0.733</td>
<td>1.35</td>
</tr>
<tr>
<td>IMF₀</td>
<td>60.6</td>
<td>118.6</td>
<td>-0.729</td>
<td>1.31</td>
</tr>
<tr>
<td>IMF₁, IMF₁₋₁</td>
<td>59.7</td>
<td>104.0</td>
<td>-0.701</td>
<td>1.25</td>
</tr>
<tr>
<td>IMF₂, IMF₁₋₁</td>
<td>60.6</td>
<td>117.7</td>
<td>-0.639</td>
<td>1.11</td>
</tr>
<tr>
<td>IMF₃, IMF₂₋₁</td>
<td>52.9</td>
<td>105.0</td>
<td>-0.577</td>
<td>0.87</td>
</tr>
<tr>
<td>IMF₄, IMF₃₋₁</td>
<td></td>
<td></td>
<td>-0.409</td>
<td>0.65</td>
</tr>
<tr>
<td>IMF₅, IMF₄₋₁</td>
<td></td>
<td></td>
<td>-0.548</td>
<td>1.05</td>
</tr>
<tr>
<td>Best estimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 11. As in Fig. 6, but with different colored lines to show the resulting spectra after filtering successive high- and low-frequency IMFs. As might be expected, the spectra decrease with each IMF removed, so, looking at the high-frequency portion of the spectra, the green line closest to the original spectra represents the removal of the highest-frequency IMF, the blue line represents the removal of the two highest-frequency IMFs, and the next green line represents the removal of the three highest-frequency IMFs. The same is true of the low-frequency portion of the spectra. Unlike Fig. 6, data presented here are not normalized, so that all spectra are superimposed on each other.
due to the aircraft movements and not to the atmospheric dynamics. As these disturbances have a significant impact on the turbulent fluxes, this is a very important contribution of this method. However, similar to any other data-processing technique, there is some subjectivity in the process of eliminating some of the IMFs and this is the most obvious weakness of this method. As we continue using it in processing the HOP data, we anticipate learning more about its benefits and flaws, especially in determining the IMFs that can be explained physically. Thus, in the near future, we anticipate eliminating much of the subjectivity applied in the present study.

Acknowledgments. Funds to cover the graduate fellowships for the first and second authors were provided by NSF Grant 0623182 and by NASA Grants NNX07AO54H and NNX08AR54G, respectively. The flights conducted in CLASIC were funded by NASA under Grant NNX07AJ60G.

REFERENCES


