Data archive strategy for computing the long-term means of nonlinear functions in geophysical problems

Ben-Jei Tsuang*

Department of Environmental Engineering, National Chung-Hsing University, 250 Kuokang Rd., Taichung 402, Taiwan

ABSTRACT: An algorithm for computing the temporal mean of a nonlinear function using the temporal means, covariances and higher order statistical moments of the variables involved in the function is revisited. Furthermore, a pyramidal algorithm is derived, which hierarchically stores the statistical moments of a longer interval of a variable from those of its shorter subintervals. The 2 methodologies together presented here show a systematic way of data storage and show that the long-term mean of a nonlinear process can be analyzed by decomposing it into various shorter sub-time scales such as diurnal and seasonal cycles. For example, the long-term mean of horizontal moisture flux can be decomposed into the product of the means of wind speed and humidity observations, plus the covariance of daily means of the 2 variables, and plus the mean of the daily covariances of the 2 variables on each day, where the 3 mean values and the covariance are suggested for storage. The results are exactly the same as those directly calculated from their hourly data. Since only 4 statistical moments are needed, significant data reduction for data distribution can be achieved. The error associated with the data archive replica has been discussed for a highly nonlinear function of which the statistical moments of its variables are only available up to a finite order. A case study using 41 yr of data taken on an urbanizing site on a subtropical island is illustrated.

KEY WORDS: Data processing · Nonlinear · Transient and time domain · Seasonal cycle · Diurnal cycle · Taichung

1. GEOPHYSICAL PROBLEMS

The climatic variations of air temperature, precipitation and runoff have been studied extensively over the past 3 decades using the wavelet method as well as other statistical methods (e.g. Murphy & Katz 1985, Hansen 1987). However, there has been little study on the long-term variation of energy budget components (Hsu & Liu 2003) using observed data. One of the difficulties is that few direct measurements of these components over long periods of time are available; therefore, they must be derived from other quantities (prognostic variables). For example, the wind kinetic energy per unit mass (KE) can be derived from wind speed s using the equation $KE = s^2/2$; the black body radiation of air R is determined from air temperature T in Kelvin using the equation $R = \sigma T^4$, where σ is the

Stephan-Boltzmann constant; the advected moisture flux direction is determined using the equation $u_j q$, where u_j is the wind vector and q is the mixing ratio. Notably, all of these equations ($s^2/2$, σT^4 and $u_j q$) are nonlinear. In this study a data archive strategy is proposed for the prognostic variables commonly used in the conservation equations of mass, momentum and energy for computing various nonlinear geophysical properties. The prognostic variables considered here include wind vector (u_j), temperature (T) and mixing ratio (q). Note that wind speed $s \equiv \sqrt{u_j^2}$.

An accurate determination of the mean value of a nonlinear function over the time domain must take into account the time variation of its variables. For example, to accurately determine the mean of the kinetic energy ($\overline{KE}$), over a period, theoretically we would have to use continuous wind speed data s from that

period for substitution into the equation. However, in actuality, the prognostic data are sampled with certain frequencies and then averaged over and stored at certain intervals. The bulk of the data is available only in hourly, daily or monthly intervals and not continuously. In particular, most climatological data are only available as monthly averages so as to be more manageable for storage or data distribution (e.g. Xie & Arkin 1997, New et al. 2000, Kistler et al. 2001). Hence, there is a discrepancy between a true value and that calculated using 'discretized' hourly, daily or monthly mean prognostic data. For example, we usually calculate KE by the equation $\left(\sum_{i=1}^{n} s_i^2/2\right)\!/n$, where s_i is the discretized data of a certain interval of wind speed, and n is the number of data over the averaging period. This discrepancy needs to be quantified.

Instead of calculating the mean of a nonlinear function using a continuous data set of each variable, this study revisits the delta method (e.g. Bishop et al. 1975, Kotz et al. 1988) for the calculation. Furthermore, for the convenience of data processing, a pyramidal algorithm is derived, by which we can determine the statistical moments of a monthly interval of a variable from those of its shorter daily intervals. However, the error can, though not always, be significant for a highly nonlinear function of which the statistical moments are only available up to a finite order. An error analysis is conducted to quantify the truncational errors. A case study using 41 yr of data taken on a subtropical island is illustrated.

2. DELTA METHOD

The delta method (also called the statistical differentials method) is used in this study. This is a method of approximating the mean of a function of random variables. It consists essentially of expanding the function as a Taylor series about the means of the random variables and then taking expected values, term by term, using only a few terms in the expansion.

Consider a nonlinear function $p(x_1, \ldots, x_m)$. Using the Taylor series expansion on the means $(\overline{x}_1, \ldots, \overline{x}_m)$ of the variables, the nonlinear function can be determined (Bishop et al. 1975, Kotz et al. 1988) as:

$$p(x_1,\ldots,x_m) = p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right) + \sum_{r=1}^{m}(x_r - \overline{x}_r)\frac{\partial p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right)}{\partial x_r}$$
$$+\frac{1}{2!}\left(\sum_{r=1}^{m}(x_r - \overline{x}_r)\frac{\partial}{\partial x_r}\right)^{2} p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right) + \ldots \tag{1}$$

where $\left(\sum_{r=1}^{m}(x_r - \overline{x}_r)\frac{\partial}{\partial x_r}\right)^{2}$ is to be expanded as an oper-

ator, so that the third term on the right-hand side of Eq. (1) is:

$$\frac{1}{2}\left(\sum_{r=1}^{m}(x_r - \overline{x}_r)^2 \frac{\partial^2 p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right)}{\partial x_r^{\,2}}\right.$$
$$\left.+2\sum_{r=1}^{m-1}\sum_{s=r+1}^{m}(x_r - \overline{x}_r)(x_s - \overline{x}_s)\frac{\partial^2 p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right)}{\partial x_r \partial x_s}\right) \tag{2}$$

Taking expected values and truncating the high order terms shown in Eq. (1), we obtain:

$$\overline{p(x_1,\ldots,x_m)} \approx p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right) + \frac{1}{2}\left(\sum_{r=1}^{m}\mu_2(x_r)\frac{\partial^2 p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right)}{\partial x_r^{\,2}}\right.$$
$$\left.+2\sum_{r=1}^{m-1}\sum_{s=r+1}^{m}\mathrm{cov}(x_r,x_s)\frac{\partial^2 p\!\left(\overline{x}_1,\ldots,\overline{x}_m\right)}{\partial x_r \partial x_s}\right) \tag{3}$$

where $\mu_2(x_r)$ is the variance of x_r and $\mathrm{cov}(x_r, x_s)$ is the covariance between x_r and x_s. Note that the second term of Eq. (1) vanishes because $\overline{(x_r - \overline{x}_r)} = (\overline{x}_r - \overline{x}_r) = 0$. Of course, higher-order terms can be included as desired, though the expressions rapidly become more complex. It should be noted that the approach is purely formal. For some values of the x's the series may not even converge. The smaller the deviations $(x_r - \overline{x}_r)$ the better the approximation to be expected, broadly speaking. A more detailed error analysis will be presented later in this paper.

In the special case of a single variable x with expected value $\overline{x}$ and central moments $\{\mu_k\}$ (with $\mu_k \equiv \overline{(x - \overline{x})^k}$):

$$\overline{p(x)} = p(\overline{x}) + \frac{1}{2}\frac{\partial^2 p(\overline{x})}{\partial x^2}\mu_2(x) + \frac{1}{3!}\frac{\partial^3 p(\overline{x})}{\partial x^3}\mu_3(x)$$
$$+\frac{1}{4!}\frac{\partial^4 p(\overline{x})}{\partial x^4}\mu_4(x) + \sum_{k=5}\frac{1}{k!}\frac{\partial^k p(\overline{x})}{\partial x^k}\mu_k(x) \tag{4}$$

Eqs. (3) & (4) show that the mean of a nonlinear function can be determined by the means, the covariances and the higher statistical moments of its variables. Hence, storing these statistical moments of variables is suggested.

3. DATA STORAGE

In actual settings, we have little control over the frequency of data storage. We do not know prior to compiling data whether we will be computing daily mean value, monthly mean value or annual mean value of some nonlinear functions of observational data. Fortunately, the statistical moments (e.g. $\overline{x}$, μ_2, μ_n) of a longer interval (monthly) can be determined from those ($\overline{x}_j$, μ_{2j}, μ_{nj}) of its shorter subintervals j's (daily) of variable x as:

$$\overline{x} = \overline{\overline{x_{ji}}} = \overline{\overline{x}_j} \tag{5}$$

$$\mu_n \equiv \overline{\left(x'_{ji}\right)^n} = \overline{\left(x_{ji} - \overline{x_{ji}}\right)^n}$$

$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\left(x_{ji} - \overline{x_{ji}}\right)^n$$

$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\left(x_{ji} - \overline{x_j} + \overline{x_j} - \overline{x_{ji}}\right)^n$$

$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\sum_{k=0}^{n}C_k^n\left(x_{ji} - \overline{x_j}\right)^{n-k}\left(\overline{x_j} - \overline{x_{ji}}\right)^k$$

$$= \sum_{k=0}^{n}C_k^n\left(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\left(x_{ji} - \overline{x_j}\right)^{n-k}\left(\overline{x_j} - \overline{x_{ji}}\right)^k\right)$$

$$= \sum_{k=0}^{n}C_k^n\left(\frac{1}{m}\sum_{j=1}^{m}\overline{\left(x_{ji} - \overline{x_j}\right)^{n-k}}\left(\overline{x_j} - \overline{x_{ji}}\right)^k\right)$$

$$= \sum_{k=0}^{n}C_k^n\overline{\overline{\left(x_{ji} - \overline{x_j}\right)^{n-k}}\left(\overline{x_j} - \overline{x_{ji}}\right)^k}$$

$$= \sum_{k=0}^{n}C_k^n\overline{\mu_{(n-k)j}\left(\overline{x_j} - \overline{x_{ji}}\right)^k} \tag{6}$$

where x_{ji} is a data point at the subinterval (day) j time (hour) i; $x'_{ji} \equiv x_{ji} - \overline{x_{ji}}$; l is the length of the subinterval j; $m \times l$ is the total length of the longer interval. $C_k^n \equiv \dfrac{n!}{k!(n-k)!}$; and $\mu_{(n-k)j}$ is the $(n-k)$th order statistical moment of a subinterval j $(= \overline{\left(x_{ji} - \overline{x_j}\right)^{(n-k)}})$. The explicit forms of the relationship for various statistical moments are listed in Table 1.

Similarly, the covariance $\mathrm{cov}(x_r, x_s)$ between variables x_r and x_s of a longer interval can be determined from those $(\overline{x_{rj}}, \overline{x_{sj}}, \mathrm{cov}_j(x_r, x_s))$ of its shorter subintervals j's as:

$$\mathrm{cov}(x_r, x_s) \equiv \overline{\left(x_r - \overline{x_r}\right)\left(x_s - \overline{x_s}\right)}$$

$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\left(x_{rji} - \overline{x_r}\right)\left(x_{sji} - \overline{x_s}\right)$$

$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\left[\left(x_{rji} - \overline{x_{rj}}\right) + \left(\overline{x_{rj}} - \overline{x_r}\right)\right]\left[\left(x_{sji} - \overline{x_{sj}}\right) + \left(\overline{x_{sj}} - \overline{x_s}\right)\right]$$

$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{l}\sum_{i=1}^{l}\left[\left(x_{rji} - \overline{x_{rj}}\right)\left(x_{sji} - \overline{x_{sj}}\right) + \left(x_{rji} - \overline{x_{rj}}\right)\left(\overline{x_{sj}} - \overline{x_s}\right) + \left(\overline{x_{rj}} - \overline{x_r}\right)\left(x_{sji} - \overline{x_{sj}}\right) + \left(\overline{x_{rj}} - \overline{x_r}\right)\left(\overline{x_{sj}} - \overline{x_s}\right)\right]$$

$$= \frac{1}{m}\sum_{j=1}^{m}\left[\mathrm{cov}_j(x_r, x_s) + 0 + 0 + \left(\overline{x_{rj}} - \overline{x_r}\right)\left(\overline{x_{sj}} - \overline{x_s}\right)\right]$$

$$= \overline{\mathrm{cov}_j(x_r, x_s)} + \mathrm{cov}\left(\overline{x_{rj}}, \overline{x_{sj}}\right) \tag{7}$$

Hence, using Eqs (6) & (7), the statistical moments at different time scales can be determined hierarchically. That is, if we only store the daily means $\overline{x_j}$ and the daily variances μ_{2j} of the variable each month, then they can be used for determining their monthly statistical moments. These monthly statistical moments can be further used for determining their annual statistical moments. Fig. 1 shows such a procedure. We call this a *pyramidal algorithm*, borrowing a term from wavelet analysis (Mallat 1989).

Table 1. Relationship between the statistical moments ($\overline{x_{ji}}$, μ_2, μ_3, μ_4) and covariance ($\mathrm{cov}(x_r, x_s)$) of a longer interval and the statistical moments ($\overline{x_j}$, μ_{2j}, μ_{3j}, μ_{4j}) and covariance ($\mathrm{cov}_j(x_r, x_s)$) of its shorter subintervals j's

Order of statistical moment	Form
0	$\overline{x} = \overline{x_{ji}} = \overline{\overline{x_j}}$
1	$\overline{x'_{ji}} = \overline{\left(\overline{x'_{ij}}\right)_j} + \overline{\left(\overline{x_j} - \overline{x_{ji}}\right)} = 0$
2	$\mu_2 = \overline{\mu_{2j}} + \mu_2\left(\overline{x_j}\right)$
3	$\mu_3 = \overline{\mu_{3j}} + 3 \cdot \overline{\mu_{2j}\left(\overline{x_j} - \overline{x_{ji}}\right)} + \mu_3\left(\overline{x_j}\right)$
4	$\mu_4 = \overline{\mu_{4j}} + 4 \cdot \overline{\mu_{3j}\left(\overline{x_j} - \overline{x_{ji}}\right)} + 6 \cdot \overline{\mu_{2j}\left(\overline{x_j} - \overline{x_{ji}}\right)^2} + \mu_4\left(\overline{x_j}\right)$
n	$\mu_n = \overline{\left(x'_{ji}\right)^n} = \sum_{k=0}^{n}C_k^n\overline{\left(x_{ji} - \overline{x_j}\right)^{n-k}\left(\overline{x_j} - \overline{x_{ji}}\right)^k} = \sum_{k=0}^{n}C_k^n\overline{\mu_{(n-k)j}\left(\overline{x_j} - \overline{x_{ji}}\right)^k}$
Covariance	$\mathrm{cov}(x_r, x_s) = \overline{\mathrm{cov}_j(x_r, x_s)} + \mathrm{cov}\left(\overline{x_{rj}}, \overline{x_{sj}}\right)$

Note that μ_2 is the variance of the longer interval ($= \overline{x'^2_{ji}}$), μ_3 is the skewness ($= \overline{x'^3_{ji}}$) and μ_4 is the kurtosis ($= \overline{x'^4_{ji}}$). $\mu_{(n-k)j}$ is the $(n-k)$th order statistical moment of a subinterval j ($= \overline{\left(x_{ji} - \overline{x_j}\right)^{(n-k)}}$), and $\mu_k\left(\overline{x_j}\right)$ is the kth order statistical moment for the subinterval means $\overline{x_j}$ of the longer interval ($= \overline{\left(\overline{x_j} - \overline{x_{ij}}\right)^k}$). Note that $\mu_{1j} = \overline{\left(x_{ji} - \overline{x_j}\right)} = 0$
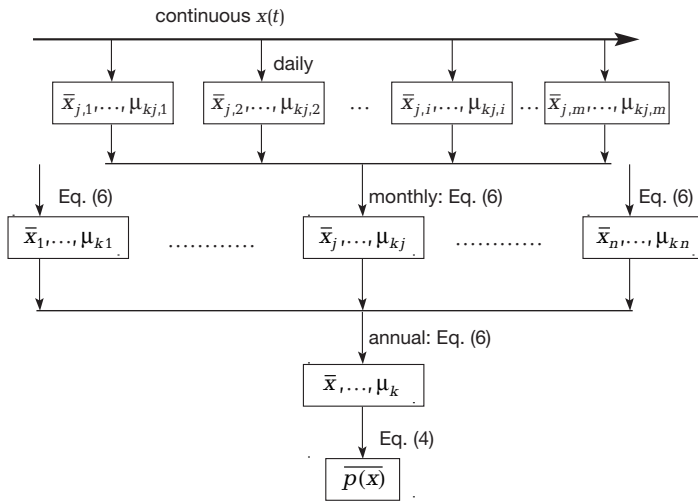
Fig. 1. Pyramidal algorithm of sampling, storage and processing data to determine the annual mean of a nonlinear process of the kth order of accuracy

4. RESULTS AND TIME SCALE DECOMPOSITION

Based on Eq. (4), the mean values of wind kinetic energy $\overline{KE}$ and black body radiation $\overline{R}$ can be determined as follows:

$$\overline{KE} = \frac{1}{2}\overline{s^2} = \frac{1}{2}(\bar{s}^2 + \mu_2(s)) \qquad (8)$$

and

$$\begin{aligned} \overline{R} &= \sigma\overline{T^4} \\ &= \sigma\overline{T}^4 + 6\sigma\overline{T}^2 \cdot \mu_2(T) + 4\sigma\overline{T}\cdot \mu_3(T) + \sigma\cdot\mu_4(T) \end{aligned} \qquad (9)$$

Based on Eq. (3), the mean value of the advected moisture flux in the x direction can be determined as follows (Boer 2000):

$$\overline{uq} = \bar{u}\cdot\bar{q} + \text{cov}(u,q) \qquad (10)$$

Eqs. (8) to (10) show that if using only the mean wind speed $\bar{s}$ to calculate the mean of $\overline{KE}$ by $0.5\bar{s}^2$, the truncational error is $0.5\mu_2(s)$. If using only the mean temperature $\bar{T}$ to calculate the mean of $\overline{R}$ by $\sigma\overline{T}^4$, the truncational error is $6\sigma\overline{T}^2\mu_2(T) + 4\sigma\overline{T}\mu_3(T) + \sigma\mu_4(T)$. If using only the mean wind component in the x direction ($\bar{u}$) and mean mixing ratio ($\bar{q}$) to calculate the mean of the moisture flux in the x direction $\overline{uq}$, the truncational error is $\text{cov}(u,q)$. Note that $0.5\mu_2(s)$ is also known as the turbulent kinetic energy when the averaging period is around 1 h. These errors shows that using only the mean of the variable to calculate the mean of a nonlinear function, the error has the first order of accuracy because the truncational error is of order $O(x'^2)$.

Furthermore, the long-term mean of a nonlinear function can be decomposed into various time components, such as annual component, daily component and hourly component. The annual component is a function of the annual means of the involved variables only, the daily component is a function of the daily means of the involved variables, and the hourly component is a function of the hourly means of the involved variables. For example, the annual mean value of the moisture flux can be decomposed according to Eqs. (7) & (10) as:

$$\overline{uq} = \overline{uq}_0 + \overline{uq}_d + \overline{uq}_h \qquad (11)$$

where $\overline{uq}_0$, $\overline{uq}_d$ and $\overline{uq}_h$ are advected moisture flux components of mean component, daily variation component and hourly variation component in the x direction. They are defined as:

$$\overline{uq}_0 \equiv \bar{u}\cdot\bar{q} \qquad (12)$$

$$\overline{uq}_d \equiv \text{cov}(\bar{u}_j,\bar{q}_j) \qquad (13)$$

$$\overline{uq}_h \equiv \overline{\text{cov}_j(u,q)} \qquad (14)$$

where subscript j indicates the data date.

5. CASE STUDY

In this section, we examine the applicability of the derived equation using hourly data measured at Taichung Meteorological Station (24°09′N, 120°41′E), operated by the Central Weather Bureau of Taiwan. Data include wind vector (u_j) and mixing ratio (q) for the period between 1961 and 2001. Note that as only hourly data are available, components of data with a frequency higher than 1 h^{-1} cannot be obtained. First, daily statistical moments are computed according to their definitions form the hourly data. Then, monthly, annual and 41 yr statistical moments are computed hierarchically according to the above pyramidal algorithm (Fig. 1). The monthly means of u_j and q, the monthly covariances between (u_j, q), and the monthly mean of daily covariances between (u_j, q) are stored for further analysis.

Fig. 2 shows the calculations of the annual and the monthly composite values of advected moisture flux in the x direction $\overline{uq}$ from 1961 to 2001, as determined according to Eq. (11). The monthly 41 yr composite means show that the mean component had a peak in the summer, the daily component had a peak in the spring, and the hourly component had a peak in the summer. The annual means show that the annual flux decreased by 48.7% (0.002 m kg s^{-1} kg^{-1}) during these 41 yr, whereas the mean component decreased by 0.00201 m kg s^{-1} kg^{-1}. The ratios of the 41 yr means among the mean component, daily variation component and hourly variation component were 82%:15%:3%, i.e. using only the annual mean u and q to calculate the annual flux, the bias is 18%. In addition, it shows that more flux resided in the daily varia-
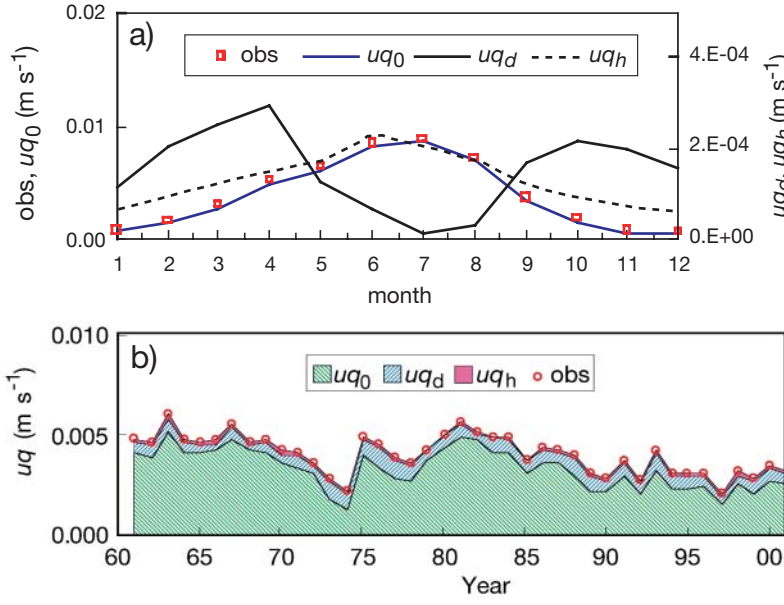
Fig. 2. (a) Composite monthly and (b) annual advected moisture flux uq from 1961 to 2001 in Taichung, Taiwan, where uq_0, uq_d and uq_h are the fluxes of mean component, daily variation component and hourly variation component, respectively, and *obs* is the flux calculated directly from hourly data

moments of the variables are available such as for black body radiation, wind kinetic energy and advected moisture flux functions. However, for a non-polynomial nonlinear function such as the saturated vapor pressure function (Richards 1971) or function $f(x, y) = \frac{y}{x}$, or for a highly nonlinear polynomial function, high order terms have to be truncated according to the Taylor series expansion if the statistical moments are only available up to a finite order. If the statistical moments are only available up to kth order, the truncational error ε of a single variable function can be determined by observing Eq. (4) as:

$$\varepsilon = \left| \sum_{i=k+1} \frac{1}{i!} \frac{\partial^i p(\bar{x})}{\partial x^i} \overline{(x')^i} \right| \qquad (15)$$

For simplicity let us inspect a nonlinear function with a form of $p(x) = ax^n$. If the statistical moments of x are only available up to $n-1$ order, then the dimensionless error $\varepsilon^\star$, which is defined as the ratio between the truncational error and $p(\bar{x})$, can be determined according to Eq. (15) as:

$$\varepsilon^\star \equiv \left| \frac{\varepsilon}{p(\bar{x})} \right| = \left| \overline{(x'^\star)^n} \right| = |\mu_n^\star| \qquad (16)$$

where $x'^\star \equiv x'/\bar{x}$, and $|\mu_n^\star| \equiv \left| \overline{(x')^n} / (\bar{x})^n \right|$. $|\mu_n^\star|$ is the magnitude of the nth order dimensionless statistical moment. Note that $p(\bar{x})$ is the conventional choice to calculate $\overline{p(x)}$ if only the mean of x is available. This equation shows that the error increases with $|\mu_n^\star|$. Therefore, for a highly nonlinear function, for which n is huge, the error can be significant if $|\mu_n^\star| \geq 1$. Table 2 lists the dimensionless statistical moments of u, v, T and q of the study site. It shows that the magnitudes of $\mu_n^\star$ of u, v and q increase with n, but that of T decreases with n. This is because data points $x \in (0, 2\bar{x})$ exist for variables u, v, and q, but not for T. This implies that for a highly nonlinear function, the truncation error is significant for variables of u, v and q, but the error is insignificant for variable T.

tion component than in the hourly variation component, and the covariance terms $(\overline{uq_d}, \overline{uq_h})$ are positive, i.e. the fluxes of daily variation component and hourly variation component were eastward.

In addition to data-archiving, the technique described herein could also be useful for the evaluation and interpretation of changes in climate parameters. However, an implicit assumption for Fourier analysis is that the data be stationary. The techniques discussed here could be used in this context if the data are detrended prior to application of the delta method.

6. EXTREMES AND ERROR ANALYSIS

From the above examples, we have found that the delta method is superior to methods using solely the means of variables for determining their nonlinear functions. There is no bias for determining the mean of a polynomial function if the required statistical

Table 2. Statistical moments for hourly data in the period of 1961 to 2001 observed in Taichung, Taiwan. The dimensionless statistical moments are defined as $\mu_i^\star \equiv \overline{(x'/\bar{x})^i}$

	n	No. of $x \in (0, 2\bar{x})$	$\bar{x}$	$\mu_2^\star$	$\mu_3^\star$	$\mu_4^\star$	$\mu_5^\star$	$\mu_{10}^\star$	$\mu_{20}^\star$	$\mu_{50}^\star$	$\mu_{100}^\star$
u	359400	207152 (58%)	0.27 m s^{-1}	1.1E+01	5.2E+01	1.3E+03	4.5E+04	3.9E+14	5.4E+34	1.5E+95	7.6E+195
v	359400	252418 (70%)	−0.78 m s^{-1}	5.2E+00	7.8E+00	1.8E+02	4.3E+03	3.3E+12	3.2E+30	3.8E+84	5.1E+174
T	359400	0 (0%)	296 K	3.8E−04	−2.8E−06	3.5E−07	−6.8E−09	3.8E−15	2.2E−25	4.7E−54	8.1E−102
q	359397	1 (0%)	12.8 g kg^{-1}	1.0E−01	−1.7E−03	2.2E−02	−1.7E−04	4.2E−01	6.3E+04	2.2E+20	1.7E+46

7. CONCLUSION

This study recalls a method for finding the mean of a nonlinear function using the means, the higher order statistical moments, and the covariances of the function's variables. This method is more accurate than the method using only the means of the variables, which has only the first order of accuracy. As illustrated for an urbanizing site, Taichung City, Taiwan, it shows that using only the mean u and mean q for computing annual moisture flux in the x direction has an error of 18%. However, this error can be corrected by including a higher order of statistical moments for the calculation. Using the analogy, it shows that at the study site in the past 41 yr the magnitudes of the advected moisture flux has declined by 49%. This change is statistically significant and most likely due to the site being affected by global warming and urbanization effects (Tsuang et al. 1996). In contrast to the Fourier transform (Kreyszig 1993), the proposed method does not require the applicable data to be assumed stationary, i.e. the method has a benefit for data compression under climatic change.

For the purpose of climatic studies, storing not only the monthly means of the prognostic variables (u_j, T and q), but also the following statistical moments are suggested, where the monthly means of various daily statistical moments can be used to study their diurnal properties. (See Boer [2000] for a similar list of variables in various budget equations.)

- Monthly variances of u_j, T and q
- Monthly covariances among u_j and (u_i, T and q)
- Monthly means of the daily variances of u_j, T and q
- Monthly means of daily covariances among u_j and (u_i, T and q).
- Monthly skewness of T
- Monthly mean of the daily skewness of T
- Monthly kurtosis of T
- Monthly mean of the daily kurtosis of T

Following this suggestion, many of the important nonlinear functions in geophysical problems, such as the continuity equation, the momentum conservation equation, the advection-diffusion equation, the wind kinetic energy function and the black body radiation function, can be calculated accurately; and the changes of the diurnal cycle, the annual cycle and the interannual cycle of a nonlinear process can be analyzed without the need for hourly data. Currently many climatic data sets, such as the ERA-15 (Gibson et al. 1999) reanalysis and CRU data (New et al. 2000), only provide the monthly means of the prognostic variables u_j, T and q without offering their higher statistical moments. As a result, nonlinear properties being functions of the prognostic variables cannot be quantified precisely. This deficiency can be corrected by providing additional higher statistical moments, and this work can serve as a reminder for the purpose.

LITERATURE CITED

Bishop YMM, Fienberg SE, Holland PW (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, MA

Boer GJ (2000) Analysis and verification of model climate. In: Mote P, O'Neill A (eds) Numerical modeling of the global atmosphere in the climate system. Kluwer Academic Publishers, Dordrecht, p 59–82

Gibson JK, Kållberg P, Uppala S, Nomura S, Hernandez A, Serrano E (1999) ERA-15 description (version 2). ECMWF Re-Analysis project Report Series 1. European Centre for Medium-Range Weather Forecasts, Reading

Hansen J (1987) Global trends of measured surface air temperature. J Geophys Res 92(11):13345–13372

Hsu HH, Liu X (2003) Relationship between the Tibetan Plateau heating and East Asian summer monsoon rainfall. Geophys Res Lett 30(20):2066, doi: 10.1029/2003GL017909

Kistler R, Kalnay E, Collins W, Saha A and 9 others (2001) The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. Bull Am Meteorol Soc 82(2):247–267

Kotz S, Johnson NL, Read CB (1988) Statistical differentials, method of. Entry in Encyclopedia of Statistical Sciences 8, John Wiley & Sons, New York, p 646–647

Kreyszig E (1993) Advanced engineering mathematics, 7th edn. John Wiley & Sons, New York

Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal 11:674–693

Murphy AH, Katz RW (eds) (1985) Probability, statistics, and decision making in the atmospheric sciences. Westview Press, Boulder, CO

New M, Hulme M, Jones P (2000) Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 Monthly grids of terrestrial surface climate. J Clim 13(13):2217–2238

Richards JM (1971) Simple expression for the saturation vapor pressure of water in the range –50°C to 140°C. Br J Appl Phys 4:L15–L18

Tsuang BJ, Wu MC, Liu CC, Chen HH (1996) Regional climatic change in Taiwan. J Chinese Inst Environ Eng 6(2):131–150

Xie P, Arkin PA (1997) Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull Am Meteorol Soc 78:2539–2558