

处理时间序列提高计算湍流通量的精度^{*}

陈红岩

(北京大学地球物理系, 北京 100871)

胡 非 曾庆存

(中国科学院大气物理研究所大气边界层物理
和大气化学国家重点实验室, 北京 100029)

摘 要 讨论了非线性趋势对时间序列的自相关函数、积分时间尺度等的影响, 并估算了其最终对湍流通量的影响, 指出有效地去除野点、噪音和非线性趋势以及非线性趋势对时间序列的自相关函数、积分时间尺度等的影响, 可以减少其对湍流通量产生较大的虚假贡献。

关键词: 时间序列分析; 非线性趋势; 湍流通量

1 引言

为得到较准确的湍流通量, 主要应从准确的观测入手。这就要求有良好的实验场地, 性能指标稳定的仪器和电源, 合适的采样频率, 还要求仪器放置平稳且高度和间距恰当, 同时要求环境比较清洁并能避免飞虫等生物的干扰等。

然而, 实际观测到的湍流资料, 很少会是完美的。受到场地、环境、不稳定电源和仪器自身的影响, 我们得到的常常是带有野点和噪声并可能是不平稳的时间序列。虽然可以通过选择较理想的场地、稳定的仪器和电源等一些措施提高数据质量, 选择出一些比较好的时间序列, 然而还是不得不放弃一些 (或许是大部分) 来之不易的资料。有时即使进行了严格的筛选, 得到的资料也往往难如人愿, 所以有必要设计合理有效的资料处理方法, 从原始资料中分析、提取有用的信息^[1~3]。

本文具体针对计算湍流感热通量和潜热通量, 探讨大气湍流资料的去野点、去噪、稳定性分析、趋势分解及时间序列长度的选择等。资料采用 1998 年 6 月淮河实验 (HUBEX) 的近地面层通量观测, 选择的场地下垫面比较均匀, 主导风 (东风) 方向是大片的水稻田, 水稻苗高度约 20 cm, 西方有一些豆子、蔬菜等作物, 西北方向有一些低矮建筑。采样仪器采用美国 SAT-211/3K 超声风速温度仪和 M100 差分红外湿度脉动仪 (Differential Absorption Hygrometer), 仪器架设高度 4 m, 采样频率分别为 20 Hz 和 10 Hz。

2 去野点与去噪

观测资料的野点可能是由于飞虫等污染物、电源的不稳定脉冲或感应元件的不稳定工作造成的, 尤其是当环境温度过热或过于潮湿时。由于不同的仪器性能有差别, 所以

1999-06-14 收到, 1999-08-11 收到修改稿

^{*} 本文获得国家自然科学基金课题 49505065 和中国科学院“九五”重点项目 KZ952-S1-405 资助

某些物理量的野点并不一定影响其他量, 野点在每个时间序列的表现特征也有所不同。

实验数据的野点可归纳为如下几种类型: (a) 孤立单野点: 一般在时间序列中是一个很大的尖峰, 而野点邻近的数据点都远小于 (或大于) 野点值; (b) 孤立的连续野点: 该类型野点一般在时间序列中也是一个很大的尖峰, 且连续出现的几个野点的左邻和右邻的数据点都远小于 (或大于) 野点值; (c) 阵发噪音型野点: 这种野点类似噪音, 在时间序列里像是有规律地发生, 然而在每组野点内部却起伏很大, 远离真实值的程度也大小不一。这样的野点如果很多, 时间序列将会变得不可用。

对于 a、b 型野点, 可以用下面的方法滤掉: 构造一个新的时间序列 $s_n = a_n - a_{n-1}$, 在野点的左邻和右邻必各有一大的跳跃点并且方向是相反的。这样的方法不会过滤掉类似温度场斜坡结构的湍流相干结构。c 型野点不大容易去除, 但由于野点都发生在高频区, 可以用低通滤波的办法, 结合与观测资料的比较来去除。去除水平风分量测量的野点, 还要考虑风向和风速的连续变化。

时间序列的信号噪音多是由于仪器自身的性能以及电源的噪音污染造成的, 噪音的频带可能也是比较宽的。一般的滤波方法是低通滤波, 也可以考虑用小波或非线性的方法, 这里不再详述。但是数字滤波去噪的实际应用效果并不理想, 所以湍流仪器的抗噪性能就比较重要。

3 稳定性分析 (去趋势)

时间序列的稳定性是指理想的时间序列的整体平均不随时间变化, 即:

$$\frac{d\langle u(t) \rangle}{dt} = 0.$$

影响时间序列的稳定性的因素有非均匀的下垫面、风向的变化、物理量的日变化、天气尺度的变化等。时间序列的分析对稳定性的要求较高, 不稳定的时间序列对计算通量是歪曲的, 其中非线性趋势项的歪曲通量可能超过实际湍流通量的贡献。

湍流通量是指湍流涡旋对各种物理量的输送。湍流涡旋的尺度范围较宽, 但主要贡献是 1 h 以下的小尺度涡旋。湍流通量一般由基于湍流统计理论的涡度相关法计算。对于较小尺度的涡旋, 实测的湍流时间序列能够很好地多次重复再现, 从而使其统计意义显著。但对于较大尺度 (例如大于 10 min) 的涡旋, 由于实测湍流时间序列的有限长度, 因而不具有统计意义。因此, 涡度相关法计算的湍流通量常常受大尺度涡旋的不确定性的影响。为了更准确地得到较确定的小尺度涡旋湍流通量, 就要进行合理的非线性趋势滤波, 这种非线性趋势滤波可能去除了一些较大尺度涡旋的湍流通量, 但是大尺度涡旋的湍流通量本来就是难以统计的, 而湍流通量更主要是由较小尺度的湍流贡献的。另外一个进行非线性趋势滤波的理由是考虑风向的摆动, 由于实验场地常常是非均匀的 (或比较均匀), 风向的变化带来了不同方向的气流, 往往是不同方向的气团具有各不相同的统计均值, 表现在实测的时间序列上就是实测时间序列带有非线性趋势。

3.1 线性去趋

由于实测的时间序列总有日变化或中大尺度变化的趋势, 所以线性去趋势总是必须

的, 如果去掉线性趋势后的时间序列基本是稳定的, 这样的时间序列就具有较好的统计意义。

然而, 实测的时间序列往往是复杂的, 带有各种尺度的非线性趋势, 这就需要进行进一步去除非线性趋势。

3.2 非线性趋势滤波

非线性的趋势项可以用高通滤波的方法过滤, 高通滤波的截止频率和时间序列的积分时间尺度有关。假设将时间序列 X 做线性去趋后, 做如下分解: $X = X_1 + X_s$, X_1 是非线性趋势项, X_s 是去趋后的平稳时间序列。

对于通量计算:

$$\begin{aligned}\overline{XT} &= \overline{(X_1 + X_s)(T_1 + T_s)}, \\ \overline{XT} &= \overline{X_1 T_1} + \overline{X_1 T_s} + \overline{X_s T_1} + \overline{X_s T_s},\end{aligned}$$

如果趋势项尺度远大于湍流涡旋尺度, 一般 $\overline{X_1 T_s} \approx \overline{X_s T_1} \approx 0$, 所以,

$$\overline{XT} = \overline{X_1 T_1} + \overline{X_s T_s}.$$

如果存在较弱的非线性趋势项时, $\overline{X_1 T_1}$ 较小, $\overline{XT} \approx \overline{X_s T_s}$, 反之, 很强的非线性趋势项的存在会掩盖了真实的湍流通量, 从而使用涡度相关法计算的湍流通量失去意义。虽然时间序列存在着不稳定的因素, 我们仍然希望去除非线性趋势项后 $\overline{X_s T_s}$ 能近似反映通量 $\overline{XT}$ 。应该注意的是, 当非线性趋势项的贡献 $\overline{X_1 T_1}$ 较大时会影响 $\overline{X_s T_s}$ 接近 $\overline{XT}$, 因为各态历经的假设在不平稳过程中是不成立的。

4 时间序列长度

时间序列的长度和稳定性是矛盾的, 往往稳定的时间序列不会持续太长, 而准确的通量计算需要尽量考虑所有尺度的输送。

假设时间序列从 $-T/2$ 到 $T/2$, 定义:

$$\mu_1(T) = \frac{1}{T} \int_{-T/2}^{T/2} u(t) dt.$$

设整体平均为 μ_1 , 有限长的时间序列的平均值与整体平均 μ_1 偏差的标准方差为

$$\sigma_1^2(T) = \langle [\mu_1(T) - \mu_1]^2 \rangle = \langle \mu_1(T)^2 \rangle.$$

通过推导可以得到^[4]

$$\sigma_1^2(T) = \frac{2}{T} \int_0^T \left(1 - \frac{\tau}{T}\right) R(\tau) d\tau.$$

设 $\mu_2 = R(0)$ 为 u 的标准差的平方, 令自相关函数

$$\rho(\tau) = \frac{R(\tau)}{\mu_2},$$

于是,

$$\sigma_1^2(T) = \frac{2\mu_2}{T} \int_0^T \left(1 - \frac{|\tau|}{T}\right) \rho(\tau) d\tau.$$

从上式可以看出, 时间序列的长度对用有限的时间序列估算方差的误差影响很大, 另一个影响因子是自相关函数, 时间序列的不稳定对自相关函数是有较大影响的^[4]。

图 1 给出的是较平稳的实测湿度时间序列的自相关函数。图 2 则是较不平稳的湿度时间序列的自相关函数。可以看出较平稳的时间序列的自相关函数, 随时间逐渐减少, 较快地减到零点以下并稳定在小值附近; 不平稳的时间序列自相关函数, 减少到某一数值 (较大) 后, 很难衰减到零。

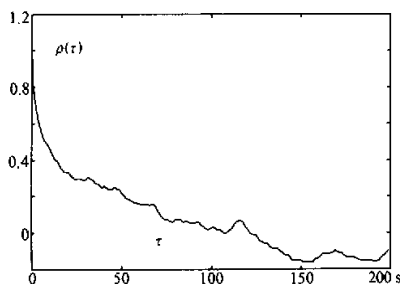


图 1 1998-06-10T17:30 湿度的自相关函数局部图
时间序列长度为 30 min

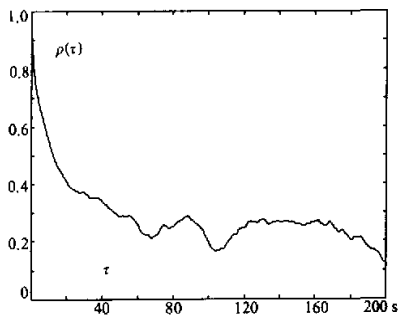


图 2 1998-06-07T12:00 湿度的自相关函数局部图
时间序列长度为 30 min

如果定义 $\tau = \int_0^\infty \rho(\tau) d\tau$ 为积分时间尺度, 当 $\rho(\tau)$ 下降较快时, 可近似地认为

$$\sigma_1^2(T) \approx 2\mu_2 \frac{\tau}{T}.$$

又设偏差 $\varepsilon = \sigma_1 / \mu_1$, 可以得到:

$$T \approx \frac{2\mu_2 \tau}{\mu_1^2 \varepsilon^2}.$$

可以看到不平稳的时间序列加大了积分时间尺度, 从而需要较长的时间序列才能得到同样精度的结果。下面将用两个实例予以说明。

取 1998 年 6 月 10 日淮河实验观测的湿度 (水汽密度) 的时间序列为例。该时间序列是较平稳的。 $\tau = 2.5$ s, $\mu_2 = 0.48$, 要求方差的误差在 10%, 即 $\sigma_1 = 0.1\mu_2$, $\mu_1 = 17.24$, $\varepsilon = \sigma_1 / \mu_1 = 0.0028$; 计算得到 $T \approx 1000$ s。也就是说在计算湿度的方差时, 为保证误差小于 10%, 时间序列长度至少要取 1000 s。若取 $T = 1800$ s, 误差减少到 $\sigma_1 / \mu_2 = 0.076$, 即 7.6%。实际上, 由于自相关函数 $\rho(\tau)$ 的特点, 积分时间尺度 τ 可能与积分 $\int_0^T \left(1 - \frac{|\tau|}{T}\right) \rho(\tau) d\tau$ 有较大差别。如上面的例子, $\int_0^T \left(1 - \frac{|\tau|}{T}\right) \rho(\tau) d\tau = 0.5$ s。所以, 真正的时间序列长度只需要 200 s 左右。

下面是不稳定时间序列的个例。采用 1998 年 6 月 7 日 12 点的实测湿度时间序列。该序列积分时间尺度 $\tau = 180$ s, $\mu_2 = 0.77$, 同样要求误差在 10%, $\sigma_1 = 0.1\mu_2$, $\mu_1 = 15.21$, $\varepsilon = \sigma_1 / \mu_1 = 0.005$; 计算得所需时间序列长度 $T = 47\,000$ s。这是由于不稳定过程非线性趋势项对自相关函数 $\rho(\tau)$ 的影响造成了 $\rho(\tau)$ 缓慢下降而使积分时间尺度增加到 $\tau = 180$ s 并且 $\int_0^T \left(1 - \frac{|\tau|}{T}\right) \rho(\tau) d\tau$ 也有 120 s 之大, 所以为保证误差 10% 时间序列长度要取几万秒, 这在实验上是不可能实现的。如果选取 3 000 s 的截止频率做高通滤波, 积分时间尺度 τ 会降至 2.3 s。因此, 对于不平稳和弱平稳的时间序列, 用较好的滤波方法滤除趋势, 减少积分时间尺度是十分必要的。

5 热通量的计算结果与讨论

超声风速温度仪测量的温度为

$$T_s = T(1 + 0.32e/p),$$

其中, e 为水气压强, p 为大气压强。

红外湿度计测量的水汽密度为

$$\rho_v = \varepsilon \frac{e}{R_d T},$$

其中, $\varepsilon = 0.622$, R_d 为干空气气体常数。

感热通量为

$$\bar{Q}_h \approx \rho c_{p,d} (\overline{w't'} + 0.84 T \overline{w'q'}),$$

其中, $c_{p,d}$ 为干空气定压比热。

潜热通量为

$$\bar{Q}_e \approx \bar{\rho} L_v \overline{w'q'}.$$

其中, L_v 为水的蒸发潜热。

图 3~6 分别为 1998 年 6 月 10 日和 1998 年 6 月 7 日的时间序列经过去趋势后计算的感热和潜热通量的日变化和趋势项对通量的虚假贡献。可以看出, 热通量有明显的日变化规律, 而且在相邻的时间段没有较大的起伏; 白天常常会出现不稳定的时间序列 (尤其是湿度), 造成的虚假贡献也会有几十 $\text{W} \cdot \text{m}^{-2}$ 的量级, 夜间总的热通量较小, 然而趋势项的虚假贡献却可能掩盖实际通量的大小。

6 结论

(1) 使用实测的湍流时间序列前必须用合理的方法去除野点和噪音, 去除野点应根据野点的特点采用相应不同的方法。

(2) 时间序列中的趋势 (多为非线性趋势) 对自相关函数影响很大, 并常常对通量的计算产生虚假贡献。

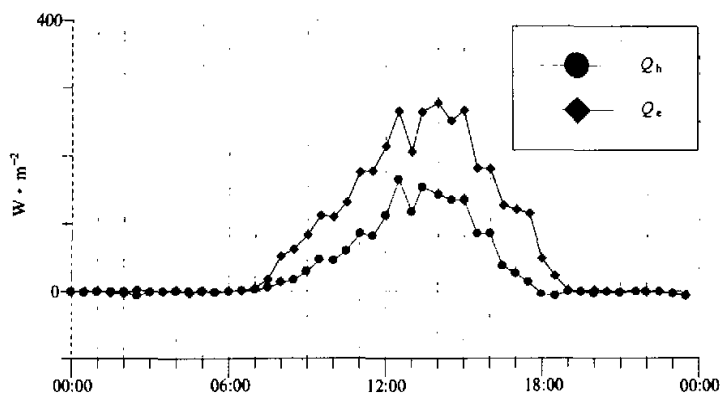


图6 1998年6月7日趋势项对感热和潜热的虚假通量

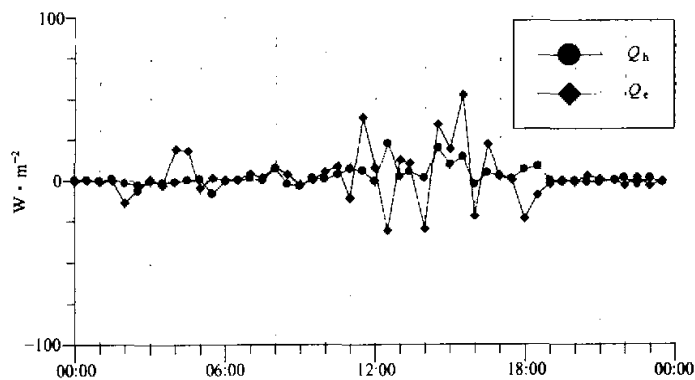


图3 1998年6月10日感热和潜热通量的日变化

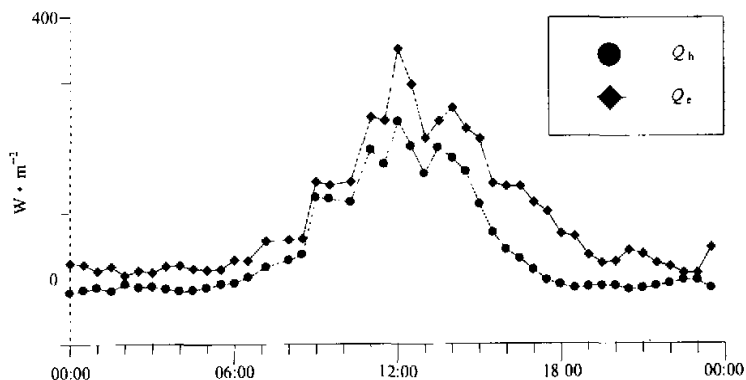


图4 1998年6月10日趋势项对感热和潜热的虚假通量

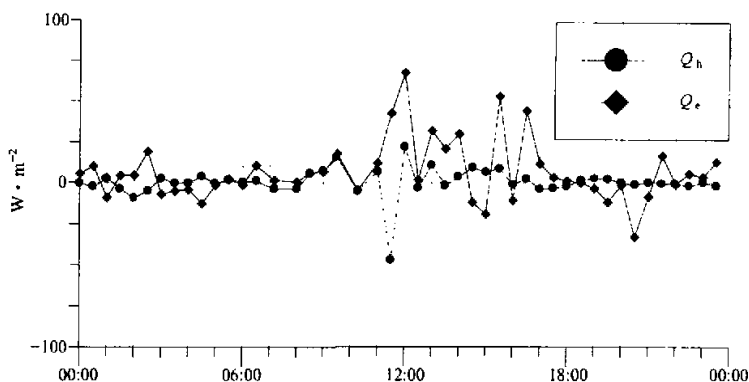


图5 1998年6月7日感热和潜热通量的日变化

(3) 计算湍流通量的时间序列长度的选择, 受积分时间尺度的影响, 合理地滤掉时间序列中的趋势项, 可以减少时间序列的积分时间尺度, 从而减少用有限长时间序列计算湍流通量的误差。

(4) 从我们的个例可以看到白天边界层趋势项对通量的虚假贡献可能达到 20%~30%, 而夜间有可能达到 100%以上。

参 考 文 献

- 1 R. B. Stull, 边界层气象学导论 (第8章), 徐静琦、杨殿荣译, 青岛: 青岛海洋大学出版社, 1991.
- 2 J. C. Kaimal, J. J. Finnigan, *Atmospheric Boundary Layer Flows -- Their Structure and Measurement* (Chapter 7), Oxford University Press, 1994.
- 3 J. L. Lumley and H. A. Panofsky, *The Structure of Atmospheric Turbulence*, Wiley Interscience, New York, 1964, 35~45.
- 4 Leif Kristensen, *Time Series Analysis -- Dealing with Imperfect Data* Riso National Laboratory, Roskilde, Denmark February, 1998.

Dealing with Imperfect Data to Improve Estimation Precision of Turbulence Flux

Chen Hongyan

(Department of Geophysics, Peking University, Beijing 100871)

Hu Fei and Zeng Qingcun

(State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry,

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029)

Abstract The observational data are seldom perfect, they are usually short noisy with wild points and nonlinear trends. In this paper, we talk about how to delete noise, wild points and filter nonlinear trends, how nonlinear trends contribute fake turbulence flux by having influence on self-relation function and integral scale.

Key words: time series analysis; nonlinear trend; turbulence flux