

近岸带波高与周期分布的核密度估计*

陈子燊 李志龙 冯砚青 李志强 常瑞莲

(中山大学水资源与环境系 广州 510275)

提要 使用双变量核密度估计方法描述近岸带波高和周期联合概率密度分布与波高、周期边缘密度分布。结果表明,核密度估计方法比通常使用的参数模式能更好地显示出具有多峰的波要素统计结构,核密度估计的波周期带宽系数能反映波浪谱的某些信息,尤其以波周期带宽和谱宽参量具有良好的线性关系。

关键词 近岸带,波高和周期联合分布,边缘密度分布,核密度估计

中图分类号 P731.22

波浪要素的统计分布是波浪过程研究的重要内容。目前,主要使用参数分布模式来描述波高、波周期分布(文圣常等,1984)。对于波高和周期联合分布,Bretschneider(1959)曾假定波高 H 与周期 T 为独立的随机变量, H 与 T 的边缘分布服从瑞利分布,建议联合概率密度函数是这两个独立变量的瑞利分布的乘积,但因 H 和 T 明显相关而失去正确性。Longuet-Higgins(1983)提出修正的波高与周期的理论联合分布,但得到的波高的边缘分布不符合瑞利分布,相关系数为零的假设与实际资料也不一致,因而上述统计性质也并非人们所期望的。Kimura(1981)应用双变量(波高与周期)的威布尔分布来表示经验联合概率密度,形状参数通过拟合观测资料来选取。按 Kimura 的结论,所有的参数都可根据谱宽参数进行估计,但实际上这些参数只有在单峰谱的条件下进行估计才是可能的。孙孚(1988)提出的波高波周期联合概率密度分布函数较 Longuet-Higgins 的理论模式有了进一步的改进。上述模式的共同点是采用有限个参数的函数关系描述波浪要素分布,能表达波浪分布的统计特性和一般变化规律。然而,这些理论分布或经验分布都仅适用于单峰谱波浪(风浪或涌浪),且近 10 年来一些利用波浪观测资料(包括涌浪和风浪构成的双峰谱在内)对波高与周期联合分布的探讨依然归为对单脊概率密度场特征的分析(陈雪英,1997;潘锦娥等,1996;黄培

基等,1995)。

但由于外海传入近岸海域的涌浪和当地风浪常形成了双峰甚至多峰谱,波高和周期常显示出很强的相关性,而且随水深减小,波浪传播过程引起的折射、反射作用、波破碎、波-波非线性作用以及当地海流的干涉作用等对波谱结构的影响加大,采用现有的理论模式或经验参数模式计算得到的波高和周期概率密度分布与实际资料差异明显,而小概率的长周期波要素对海岸工程建设、海岸泥沙运动及地形演变有重要意义。有关研究曾指出,实测波高和周期的联合分布通常与理论结果存在较大差异,而差异与偏度、谱宽度、波高和周期的相关系数关系密切,波浪破碎对波高和周期的联合分布影响较大(丁平兴等,1995)。近年来,国外研究人员尝试采用非参数密度估计方法来描述观测波浪要素分布,是值得重视与考虑的(Athanassoulis *et al.*, 2002; Ferreira *et al.*, 2002)。

1 核密度估计

核密度估计是以统计尤其是核理论为基础,每个数据点有各自的权值,而这些权值被分布在数据点各自周围的空间中,描述这些权值的分布情形可通过核函数来实现,如均匀(Uniform)核函数、依潘涅契科夫(Epanechnikov)核函数、三角(Triangle)核函数、四次(Quartic)核函数、三权(Triweight)核函数、高斯(Gauss)核函数、余弦(Cosinus)核函数等。核函数 $K(u)$ 满足以下条件:

* 国家自然科学基金资助项目,40276030 号。陈子燊,教授, E-mail:eesczs@zsu.edu.cn

收稿日期:2005-09-16,收修稿日期:2006-12-04

$\sup|K(x)x| < \infty$; $\lim_{x \rightarrow \infty} |x|K(x) = 0$; $\int K(x)dx = 1$ 。

核估计主要由带宽(窗宽或光滑)系数 h 和核密度函数 K 构成。

1.1 一维核密度函数

设随机样本 X_1, X_2, \dots, X_n 为抽自密度函数 $f(x)$ 的分布, 则 $f(x)$ 的核估计为

$$\hat{f}(x) = \frac{1}{nh_n} \sum k\left(\frac{x - X_i}{h_n}\right) \quad (1)$$

式中, $\hat{f}(x)$ 为总体密度 $f(x)$ 的一个核估计; $k(\cdot)$ 为核函数, 是一概率密度; h_n 为带宽; n 为样本容量。

1.2 多维核密度函数

把一维核密度扩展到多维核密度函数。设 X 为 d 维随机变量, X 的概率密度函数 $f(X)$ 核估计为:

$$\hat{f}(X) = \frac{1}{nh_n^d \det(S)^{1/2}} \sum_{i=1}^n K\left[\frac{(X - X_i)^T S^{-1} (X - X_i)}{h^2}\right] \quad (2)$$

式中, $X = (X_1, X_2, \dots, X_d)^T$, $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$, $i = (1, 2, \dots, n)$, d 是向量 X 的维数; S 是 X 的 $d \times d$ 维对称正定样本协方差矩阵。本文中作者采用依潘涅契科夫核作核函数, d 维的依潘涅契科夫核函数为:

$$K(X) = K(X_1 - X_{i_1}, \dots, X_d - X_{i_d}) \\ = \left(\frac{3}{4}\right)^d \frac{1}{H_1 H_2 \dots H_d} \prod_{1 \leq i \leq d} \{1 - [(X_i - X_{i_i})/H_i]^2\} \quad (3)$$

式中, H_1, H_2, \dots, H_d 是核函数在每一维上的带宽。

对无因次波高 $X(X = H/\bar{H})$ 和周期 $Y(Y = T/\bar{T})$ 联合分布, 波高的条件概率密度函数为:

$$\hat{f}(X|Y) = \frac{f(X, Y)}{\int f(X, Y) dY} = \frac{f(X, Y)}{f_X(X)} \quad (4)$$

式中, $f_X(X)$ 为波高 X 的边缘密度函数。同样可定义波周期的条件概率密度函数和边缘密度函数。

1.3 核密度估计精度

核估计既同样本有关, 又与 $K(\cdot)$ 和 H 的选取有关。在给定样本后, 核估计的精度取决于 $K(\cdot)$ 及 h 的选取是否适当。采用渐近积分均方误差准则 AMISE (Asymptotic Mean Integrated Squared Error) 进行度量:

$$\text{AMISE}(H) = n^{-1} R(K) |H|^{-1/2} + \int [\hat{f}(X) - f(X)]^2 dX \quad (5)$$

式中, H 是带宽参数矩阵; $R(K) = \int K(X)^2 dX$ 。上式右侧第一项为渐近积分方差, 第二项称为积分均方误差(MISE):

$$\text{MISE} = \int [\hat{f}(X) - f(X)]^2 dX \\ = \int \hat{f}^2(X) dX - 2E[\hat{f}(X)] + \int f^2(X) dX \quad (6)$$

MISE 由偏差和方差组成。当 $K(X)$ 固定时, 若 H 选得过大, 偏差较大, 但降低了方差, 故 $\hat{f}(X)$ 对 $f(X)$ 有较大的平滑, 使得 $f(X)$ 的某些特征被掩盖起来; 若 H 选得过小, 偏差减小了, 但增大了方差, 则 $\hat{f}(X)$ 有较大的波动。因此如何协调偏差和方差对于核估计精度十分重要, 需要在兼顾偏差性(有偏或无偏)与离散性(方差的大小)的原则下来建立估计量为“最优”的准则, 本文中作者使用光滑参数的正态尺度估计(Normal Scale Estimate of Smoothing Parameter)方法求取 AMISE 带宽系数矩阵 H (Wand *et al*, 1995):

$$H_{NS} = \left(\frac{8\sqrt{\pi}R(K)}{3\mu_2^2(K)}\right)^{\frac{1}{5}} \begin{pmatrix} n \\ d \end{pmatrix}^{-\frac{1}{5}} \hat{\sigma} \quad (7)$$

式中, $R(K) = \int K^2(X) dX$; $\mu_2(K) = \int X^2 K(X) dX$; $\hat{\sigma}$ 为 X 的均方差。

2 核密度函数估计的数值模拟

为了说明核密度估计对波高与周期联合分布的适用性, 首先对双峰参数谱(如: Ochi-Hubble 谱、Torsethaugen 谱)作数值模拟。Ochi 等(1976)利用 800 个谱的资料, 把风浪和涌浪共存的双峰谱以各自的波高、周期和形状参数构成的风浪谱和涌浪谱叠加的合成谱来表示此海况。该谱包括 6 参数, 除了有效波高外, 其余 5 个参数都可通过 H_s 表达, 实际的谱模型由一平均谱和对应于 95% 置信水平的 10 个变型组成, 因此本质上是一单参数谱。Ochi 和 Hubble 双峰谱(以下简称 O-H 谱)由下式表达:

$$S(f) = S_w(f) + S_s(f) \\ = \sum_{j=1}^2 \frac{H_{sj}^2 T_{pj}}{4\Gamma(\lambda_j)} \frac{(\lambda_j + 0.25)^{\lambda_j}}{(T_{pj} f)^{4(\lambda_j + 1)}} \exp\left[-\frac{(\lambda_j + 0.25)}{(T_{pj} f)^4}\right] \quad (8)$$

式中, $\Gamma(\cdot)$ 是伽玛函数。 $H_{sj}, T_{pj}, \lambda_j, j = 1, 2$ 分别表示风浪、涌浪的有效波高、谱峰周期和峰度因子。

另采用人工合成的三峰随机波面序列作模拟。三峰随机序列由下式表示:

$$\eta(t) = \sum_{i=1}^3 a_i \cdot \cos[f_i \cdot 2\pi \cdot t + \varepsilon_i(t)]$$

式中, $\varepsilon_i(t)$ 为在 $0-2\pi$ 区间均匀分布的随机量;
 a_i 为振幅; f_i 为峰频。

数值模拟具体步骤如下:

对给定的有效波高(对 O-H 谱), 计算得到 O-H 谱, 或计算三峰波面序列的波谱, 以此为靶谱, 采用等能量分割的方法, 对靶谱加以离散, 再作波面模拟(黎满球等, 1996), 对模拟的波面作谱估计与靶谱对比, 再对模拟的波面采用上跨零方法得到波高及其对应的周期。由双变量核密度估计得到的波高和周期联合分布。

模拟结果见图 1、图 2, 模拟频谱的峰值周期和靶谱输入值完全一致, 由谱矩计算的有效波高与输入值相差很小。由双变量核密度估计得到的波高和周期联合分布图(图 1、图 2)清楚地显示出涌浪与风浪组成的两个分量或多谱峰情况下的多脊波高和周期联合概率密度分布特征。

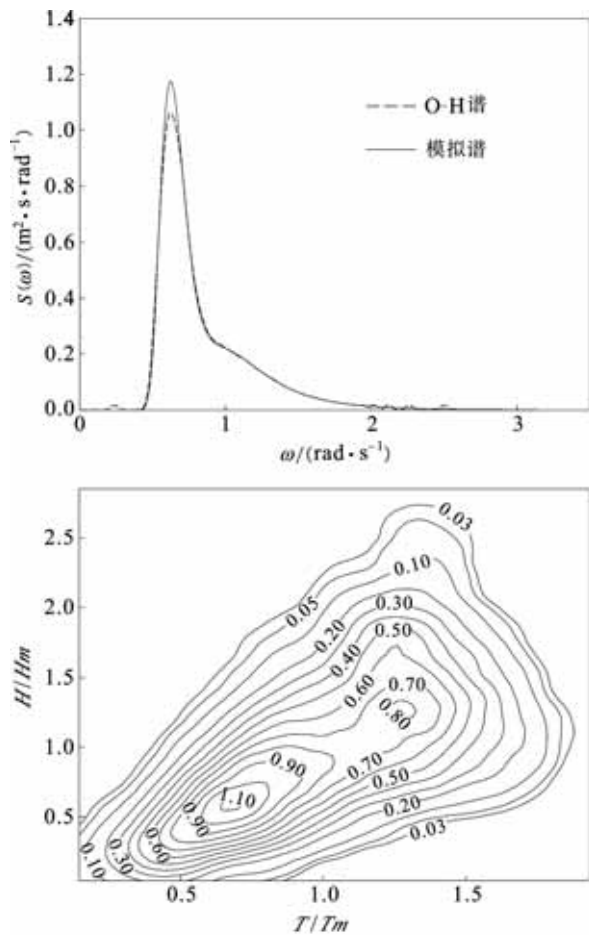


图 1 O-H 谱模拟结果($H_s=2.5\text{m}$)和周期与波高联合分布的核密度估计

Fig.1 Simulation result of Ochi-Hubble spectrum and the kernel density estimate of joint distribution of wave period and height

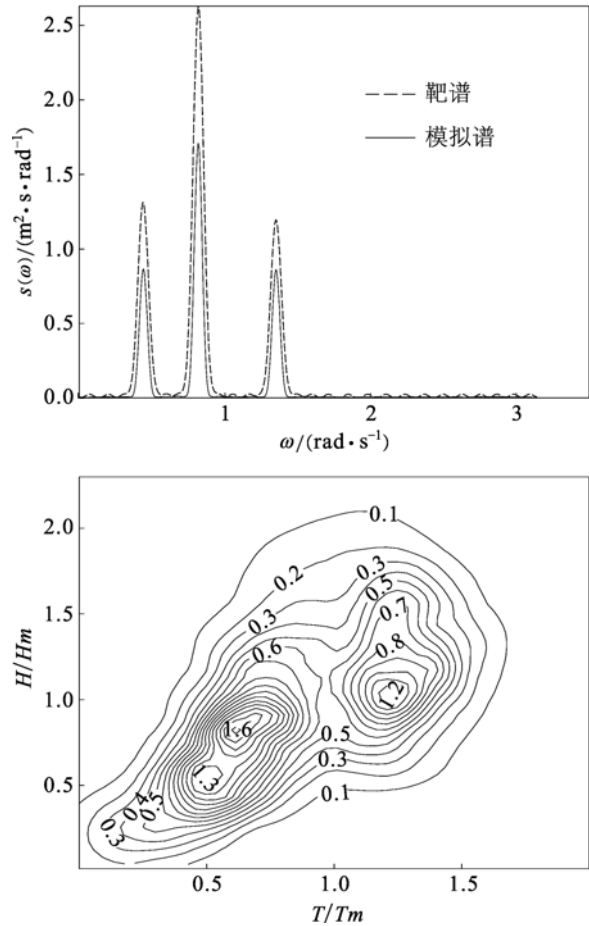


图 2 三峰波频谱模拟结果和周期与波高联合分布的核密度估计

Fig.2 Simulation of tri-peak spectrum and the kernel density estimate of joint distribution of wave period and height

3 实测资料分析

实际资料分析采用了 2003 年 8 月 6—13 日在粤东后江湾近岸带测量的部分压力波数据。测波仪分别安放在水下沙坝海侧斜坡(平均水深约 2.6m)和岸侧的凹槽内(平均水深约 1.7m)。每个测点的波浪记录数为 46 组, 各组数据长度为 4096 个, 采样频率为 2Hz。对各组记录数据处理与计算如下(Tucker, 1993): 去趋消除零漂; 计算波列偏度和峰度; 由上跨零得到波高和周期, 计算波高和周期的相关系数 R ; 计算波功率谱, 并订正为波面谱; 由谱矩 $m_p, p=0, \dots, 4$, 得到有效波高: $H_{m0} = 4\sqrt{m_0}$; 平均周期: $T_{m02} = 2\pi\sqrt{m_0/m_2}$; 谱宽度参量: $\varepsilon = \sqrt{1-m_2^2/m_0m_4}$; 谱尖度参量: $Q_p = 2/m_0^2 \int_0^\infty fs^2(f)df$; 显著波陡: $S_s = 2\pi H_{m0} / [gT_{m02}^2 \tanh(k_z d)]$; 谱峰周期: $T_p = 1/\max[s(f)]$ 。

表 1 波谱参量与波列特征值变化范围

Tab. 1 Measurements of wave spectral parameters and eigenvalue of wave trains

测点	水深(m)	偏度	峰度	R	H_{m0} (m)	T_p (s)
浅水区	2.2—3.2	0.38—1.13	2.92—5.08	0.4—0.71	0.52—1.05	8.7—15.5
碎波带	1.3—2.1	-0.09—0.9	2.78—4.51	0.47—0.8	0.37—0.79	8.4—64.0
测点	S_s	K_a	γ	ε	Q_p	
浅水区	0.008—0.015	0.09—0.34	0.69—0.79	0.4—0.58	0.98—1.52	
碎波带	0.002—0.011	0.02—0.42	0.49—0.77	0.46—0.99	0.56—1.72	

波群性参量: $Ka = \int_0^\infty |s(f) \exp(i22\pi f T_{02}) df| / m_0$;
不规则性参量: $\gamma = m_2 / \sqrt{m_0 \cdot m_4}$ 等。

以上各式中, d 为水深, k_z 为由平均周期计算的波数。表 1 为测点波特征值的计算结果。现场观测发现, 水下地形变化迅速, 观测期间, 海滩滩面坡度较大, 高潮阶段滩面引起的反射波对入射波有一定的干涉作用, 水下沙坝不断向岸推移, 并最终使得凹槽填平消亡。地形变化对波浪特性有实质性的影响, 波浪非平稳且非线性特征十分显著。波高和周期的相关性高, 谱宽度大、尖度小。峰频前后侧均可出现次级谱峰表明波-波间的非线性相互作用使得能量可向高低频两侧转移, 甚至谱峰频出现在长重力波频率范围(0.04—0.004Hz)内(图略)。

表 2 为两个测点波记录的波高、周期双变量核密度估计的带宽系数范围及组成波数。部分无维周期和波高联合概率密度分布见图 3。图 3 清楚地显示出破、碎波带波高和周期联合分布的多脊特征和高度的时间变化性。周期和波高联合概率密度偏向高低侧的概率密度脊是涌浪叠加在风浪场或组成波间能量向低频侧转移的结果有待作进一步的分析, 但这些图形真实地反映出了碎波

表 2 周期波高联合分布核估计的带宽变化范围

Tab. 2 Variation in bandwidth of kernel estimate on joint distribution of wave height and period

测点	组成波数	周期带宽	波高带宽
浅水区	227—326	0.245—0.363	0.28—0.407
碎波带	304—383	0.24—0.512	0.255—0.422

带波高和周期联合分布的概貌。

对周期和波高联合概率密度分布的核估计结果通过边缘概率密度计算分别得到波高和周期概率分布(图4), 计算结果同时与理论参数分布和经验参数分布加以对比。图4中清楚地显示出, 无论是实际数据的偏差还是与众数的符合情况, 核密度估计结果优于目前采用的参数分布模式(见图4中的瑞利波高分布、威布尔分布、广义伽玛分布、Longuet-Higgins 周期分布), 并能较好地刻画参数分布模式不能描述的多峰谱特征。

为了了解核密度估计结果和波谱或波列特性的关系, 分别对波高带宽系数 h_H 和周期带宽系数 h_T 与谱宽度参量、谱尖度参量、显著波陡、谱峰周期、波群性参量、不规则性参量、偏度、峰度、波高周期间的相关系数作线性拟合。结果显示, 波高带宽系数 h_H 和上述特征量之间的线性关系较差, 相关系数小, 一般小于 0.5。周期带

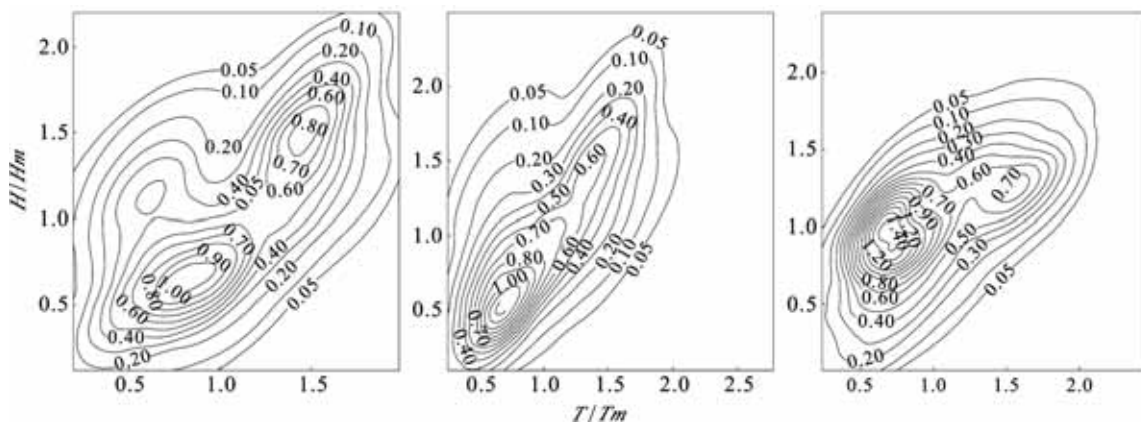


图 3 浅水区测点周期和波高联合概率核密度分布

Fig.3 Kernel density of joint distribution of wave height and period in shallow water

自左向右 3 张图的时间分别为: 2003 年 8 月 7 日 2 时; 2003 年 8 月 7 日 6 时; 2003 年 8 月 7 日 10 时

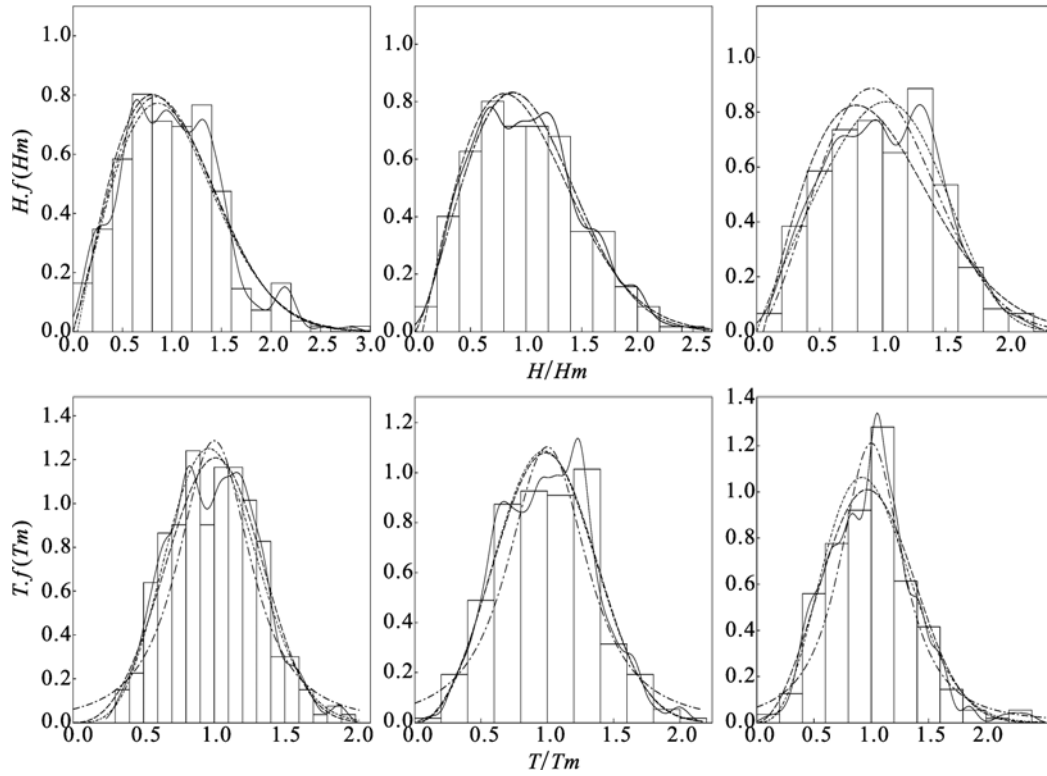


图 4 浅水区测点波高、周期分布

Fig. 4 Distributions of wave height and period in shallow water

上排图: —— 核密度分布; --- 瑞利分布; -·-·- 威布尔分布; ····· 广义伽玛分布

下排图: —— 核密度分布; --- 威布尔分布; -·-·- L-H 分布; ····· 广义伽玛分布

上、下两排图自左向右的时间分别为: 2003 年 8 月 7 日 2 时; 2003 年 8 月 7 日 6 时; 2003 年 8 月 7 日 10 时

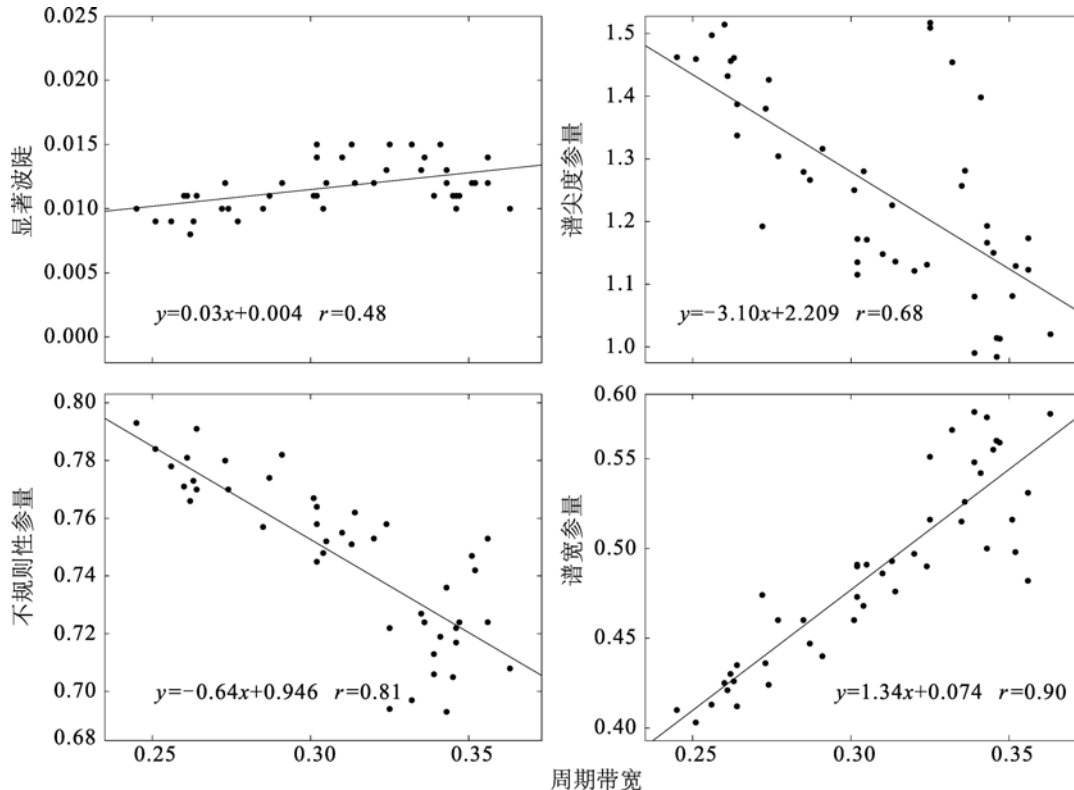


图 5 浅水区测点周期带宽与波谱参数的关系

Fig. 5 Relation between bandwidth of wave period and parameter of wave spectrum in shallow water

宽系数 h_T 和谱宽参量、不规则性参量、显著波陡、谱尖度参量相关性较高(图 5), 其中以 h_T 和谱宽参量、不规则性参量线性关系更密切。当使用不同的波浪记录数时, 带宽系数和波谱参量之间的相关性有所变化, 但波周期带宽和谱宽参量始终保持了良好的线性关系。

4 结语

核密度估计理论是一种所谓基于数据驱动的非参数统计模型, 其避开了预设模型和参数估计不确定性问题, 能在一定最优准则下, 描述非平稳与非线性时间序列的一些复杂关系, 无论估计的概率密度函数是单峰或多峰都能适用, 而且适用于高维数据结构分析, 因此核估计方法具有广泛的适应性。本文中作者对碎波带波要素概率密度分布的核估计说明, 此方法适用于对整个近岸带海浪要素密度分布的描述。

不过, 从某种意义上而言, 作为核密度估计结果的带宽系数即是参数, 是所选用的核函数的尺度参数。本文中计算过程表明: 使用不同的带宽估计方法时, 联合概率密度分布有所差异; 带宽系数能提供波浪谱的信息, 但这方面还应加以进一步的研究分析。

对观测的近岸碎波带波浪初步分析结果可见, 核密度估计应用于波高和周期概率密度分布显示出了良好的性质, 由于实际海浪的多向性, 可进一步使用三变量核密度估计方法对波高、周期和波向的联合概率密度分布加以描述。在实际应用方面, 建议采用核密度估计来表示波候的统计分布特性, 比较目前普遍使用的经验概率密度统计分布结果将更具科学性。

参 考 文 献

- 丁平兴, 孔亚珍, 孙孚等, 1995. 风浪统计性质的实验研究 II. 波高和周期的联合统计分布. 海洋学报, 17(5): 1—9
- 文圣常, 余宙文, 1984. 海浪理论与计算原理. 北京: 科学出版社, 617—618
- 孙孚, 1988. 海浪周期与波高的联合分布. 海洋学报, 10(1):10—15
- 陈雪英, 胡泽建, 1997. 双峰谱型海浪波高与周期的联合分布. 黄渤海海洋, 15(4):8—13
- 黄培基, 陈雪英, 胡泽建, 1995. 双峰谱型海浪的统计性质. 海洋学报, 17(6):1—8
- 潘锦娥, 陈志宏, 1996. 海浪波高与周期联合概率密度分布的研究. 海洋通报, 15(3): 1—13
- 黎满球, 陈俊昌, 1996. 由模拟波面统计分析波高周期联合分布. 热带海洋, 15(3): 36—43
- Athanassoulis G A, Belibassakis K A, 2002. Probabilistic description of metocean parameters by means of kernel density models 1. Theoretical background and first results. Applied Ocean Research, 24: 1—20
- Bretschneider C L, 1959. Wave variability and wave spectra for wind-generated gravity waves. U.S. Army Corps of Engineers, Beach Erosion Board, Tech Memo, 118: 1—192
- Ferreira J A, Soares C G, 2002. Modelling bivariate distributions of significant wave height and mean wave period. Applied Ocean Research, 24: 31—45
- Kimura A, 1981. Joint distribution of wave heights and periods of random sea waves. Coastal Engineering in Japan, 24: 77—92
- Longuet-Higgins M S, 1983. On the joint distribution of wave periods and amplitudes in a random wave field. Proc R Soc Lond A, 389: 241—258
- Ochi M K, Hubble E N, 1976. On six-parameter wave spectra. Proc 15th Conf Coastal Engng, 1: 301—328
- Tucker M J, 1993. Recommended standard for wave data sampling and near-real-time processing. Ocean Engineering, 20(5): 459—474
- Wand M P, Jones M C, 1995. Kernel smoothing. Chapman and Hall, 60—63

KERNEL DENSITY ESTIMATES ON DISTRIBUTION OF WAVE HEIGHT AND PERIOD IN NEARESHORE ZONE

CHEN Zi-Shen, LI Zhi-Long, FENG Yan-Qing, LI Zhi-Qiang, CHANG Rui-Lian
(*Department of Water Resources and Environment, Zhongshan University, Guangzhou, 510275*)

Abstract Statistical distribution of wave elements is important for understanding wave processes. At present, the distribution of wave period and that of wave height, or joint distribution of the both are described mainly by parametric distribution functions. For multi-peak waves that commonly seen in nearshore zones, a conventional approach would not work as it works only for single-peak cases. Other nonparametric estimates are therefore needed.

Using wave trains of numerical simulation of Ochi-Hubble spectrum of two-peak and an artificial three-peak spectrum, kernel density estimate was tested for describing the joint distribution of wave height and period. The joint distribution of wave-period-and-height and corresponding marginal probability density were simulated with two-variable kernel density estimate for the case of nearshore coastal zone east of Guangdong, China.

The results indicate that a multi-peak statistical structure could be described by kernel density estimate, but by conventional parametric models. Moreover, some features of wave spectrum were found well related to the bandwidth of wave period.

Key words Nearshore zone, Distribution of wave height and period, Kernel density estimate, Marginal density distribution