

脉动风速随机傅里叶谱模型的参数聚类分析

洪旭¹, 彭勇波², 李杰^{1,2}

(1. 同济大学 土木工程学院, 上海 200092; 2. 同济大学 土木工程防灾国家重点实验室, 上海 200092)

摘要: 脉动风速随机 Fourier 谱模型为脉动风速模拟提供了一条基于物理的建模途径. 为了实现对这一模型基本参数概率分布的合理建模, 阐述了模型基本参数的物理含义, 利用实测风速记录识别了基本参数的概率分布. 分析发现, 分界波数的统计分布出现典型的双峰形态; 由原始识别结果的统计分布导出的功率谱与 Kaimal 谱在中、低频段有较大差别. 深入研究表明, 分界波数有显著的聚类特征. 基于此, 采用 k -means 函数聚类方法分析了分界波数的数据结构, 指出上述异常现象与参数的聚类结果密切相关. 根据聚类结果, 合理地筛选实测风速记录, 进行了分界波数的建模.

关键词: 随机傅里叶谱模型; 聚类分析; 分界波数; 剪切波速; 随机建模

中图分类号: TU973.213

文献标志码: A

Cluster Analysis of Parameter of Stochastic Fourier Spectrum for Fluctuating Wind Speeds

HONG Xu¹, PENG Yongbo², LI Jie^{1,2}

(1. College of Civil Engineering, Tongji University, Shanghai 200092, China; 2. State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai 200092, China)

Abstract: The stochastic Fourier spectrum provides a physical based perspective for fluctuating wind simulation. To properly model the probability density function of the elemental variables, the basic physical meaning of the elemental variables is elaborated and the statistical distributions of them are identified from the measurements. It is revealed that the statistical distribution of the cut-off wave number displays an obvious bimodal pattern; the power spectrum density derived from the fitted distribution of the original identifications deviates from the Kaimal spectrum evidently in the low and middle frequency band. Besides, the identifications of the cut-off wave number are eminently characterized by a clustering phenomenon, based on which the

k -means cluster analysis is utilized to reveal the underlying structure of the data set. It is pointed that the cluster characteristic of the fluctuating wind speed sample is closely related to the above abnormal phenomenon. Finally, the distribution modeling for the cut-off wave number is conducted for the reasonably selected measurements based on the cluster results.

Key words: stochastic Fourier spectrum; cluster analysis; cut-off wave number; shear velocity; stochastic modeling

作为数据分析方法的一种, 聚类分析旨在将数据集集合按相似程度进行分类^[1]. 在上世纪 40 年代, Tryon 在心理学研究中最早提出了聚类分析方法^[2]; 在当代, 通过聚类分析, 互联网搜索引擎可以针对用户的查询需求将搜索结果分类分层显示^[3]; 在地球物理科学中, 聚类分析被用于影响气候的主要因素的研究^[4]; 在分子生物学中, 聚类分析在功能基因组的研究中起到重要作用^[5], 等等. 考察应用聚类分析的动机与结果不难发现, 聚类分析的主要用途在于揭示数据之间的潜在联系与内部结构, 并根据这些特征结构合理地预测其他属性. 由于数据之间特征组织方式的形成往往反映着特定物理背景的存在, 因此聚类分析能够为物理模型勾勒出大致轮廓. 显然, 这种分析方法也可以应用于风工程中.

在风工程学科诞生之初, 不少学者依据对自然风速观测记录的研究, 建立了以功率谱密度为核心的统计分析方法^[6-8]. 然而, 尽管功率谱密度模型抓住了风速随机过程的主要特征, 却并不能描述风中湍流的全部概率信息, 也难以对非平稳的风速时程做出合理的反映.

为了解决传统结构动力激励建模中概率信息难以完备描述的问题, 文献[9-11]提出了物理随机建

收稿日期: 2017-08-10

基金项目: 国家自然科学基金重点项目(51538010)

第一作者: 洪旭(1993—), 男, 博士生, 主要研究方向为结构风工程. E-mail: hongxu@tongji.edu.cn

通讯作者: 李杰(1957—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为工程可靠性理论、随机力学与混凝土损伤力学.

E-mail: lijie@tongji.edu.cn

模方法,明确指出物理规律是概率信息转移的轨道.对于脉动风速时程,文献[12-14]基于大气边界层能谱与零点演化时间概念给出了模拟脉动风速的随机 Fourier 谱模型.研究证明,该模型对实测结果的概率统计特征具有良好的刻画能力.

物理随机系统的随机性源自基本随机变量,为了合理描述脉动风速乃至结构风致响应的概率信息,需要仔细处理基本随机变量及其概率分布.应该指出:在上述随机 Fourier 谱模型中,对关键参数——分界波数的处理仍是粗糙的;已有工作的数据基础局限于个别风场观测台阵,模型对不同地域的适用性仍是需要研究的问题.有鉴于此,本文利用厦门风场实测记录,借助聚类分析方法,分析了随机 Fourier 谱中基本随机变量的分布,合理选择数据并进行建模.

1 随机 Fourier 谱模型

物理随机系统的观点认为,物理规律是随机性传播的途径^[10].如果记脉动风速沿时间 t 的变化为 $u(\Theta, t)$,对于给定的参数 Θ ,就能够获得一条风速时程 $u(\Theta, t)$.将该样本时程 $u(\Theta, t)$ 进行 Fourier 变换,可得到等价的描述方式

$$F(\Theta, n) = |F(\Theta, n)| e^{i\varphi(\Theta, n)} = \frac{1}{\sqrt{T}} \int_{-\infty}^{\infty} u(\Theta, t) e^{-i2\pi n t} dt \quad (1)$$

式中: n 为自然频率; T 为脉动风速的总时长; $|F(\Theta, n)|$ 称为幅值谱,它说明了能量在不同频率的分布; $\varphi(\Theta, n)$ 称为相位谱,它确定了不同频率处能量在时间上的分布.

由于 Θ 为随机变量,上述随机 Fourier 谱(幅值谱与相位谱)也是随机函数.因此,只要给出幅值谱与相位谱的物理模型并确定随机参数的概率信息,就能对脉动风速随机过程进行合理的反映.

实际上,充分发展的湍流由一系列不同尺度的涡旋组成.大尺度涡旋在其与主流剪切的相互作用中吸取能量;而各尺度的涡旋又不断破碎形成更小尺度的涡旋直到黏性起主导作用的 Kolmogorov 尺度,在此过程中能量逐级传递;最终在黏性的作用下湍流能量耗散为大气的内能.研究指出,在与主流相互作用强烈的含能子区和能量传递达到平衡的惯性子区中,能谱密度随波数的变化呈不同幂律关系^[15].据此,文献[12]提出了双线性 Fourier 幅值谱模型

$$|F(\Theta, k)| = \begin{cases} \sqrt{\alpha_1} \frac{u_*}{(\kappa z k_c)^{1/3}} k^{-1/2}, & k < k_c \\ \sqrt{\alpha_1} \frac{u_*}{(\kappa z)^{1/3}} k^{-5/6}, & k \geq k_c \end{cases} \quad (2)$$

式中: k 为波数; z 为高度; $\alpha_1 = 0.55$, 为 Kolmogorov 常数; $\kappa = 0.4$, 为 von Karman 常数; $|F(\Theta, k)| = \sqrt{\frac{U}{2\pi}} |F(\Theta, n)|$, 为波数谱($k = 2\pi n/U$); U 为平均风速; $\Theta = (u_*, k_c)$; u_* 为剪切波速; k_c 为分界波数,表征了含能子区与平衡子区的界限.图 1 描述了涡旋破碎与能量传递的过程.

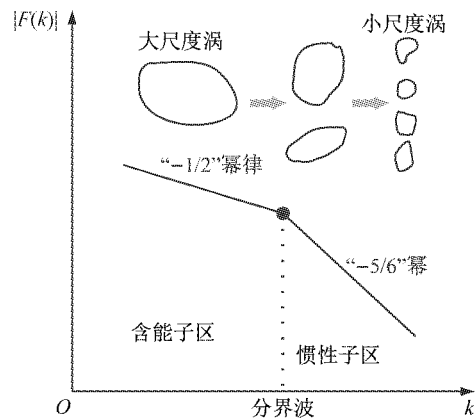


图 1 能量级串与 Fourier 幅值谱

Fig. 1 Energy cascade and amplitude of Fourier spectrum

在文献[13]中,分界波数可以通过涡量比值确定,但这一途径应用了比较粗糙的经验拟合方法,本质上是统计均值.本文的统计结果表明,分界波数具有较大的变异性,将其视为独立的随机变量是更合理的处理方式.即对于随机 Fourier 幅值谱,剪切波速 u_* 与分界波数 k_c 应分别作为独立随机变量.

在相位的研究方面,文献[14]建立了演化相位谱,它认为不同频率的涡旋振动速度不同,因而相位变化速度不同,通过合理猜想,这些涡旋具有相同的相位演化起点,并将共同起点时刻到实际观测时刻经历的演化时间定义为零点演化时间 T_e ,通过幅值谱可以得到涡旋的振动速度为

$$v(\Theta, n) = \sqrt{|F(\Theta, n)|^2 \Delta n} \quad (3)$$

式中: Δn 是频率宽度.于是,对应频率 n 的演化相位为

$$\Phi(\Theta, n, T_e) = v(\Theta, n) k(n) T_e \quad (4)$$

注意到零点演化时间在物理上表征了大气中旋涡相位的演化时间,它在时间尺度上一般较大,且受局部场地风环境影响较小,因此文献[14]的结果是合

理的。

值得指出的是,作为风场平均风的描述,风剖面对刻画风场的宏观特性十分重要.在诸多风剖面模型中,对数律的风剖面因为有一定的理论基础且与实际结果较吻合而得到广泛的应用^[16].因此,本文采用对数律作为风场平均特性的度量,即

$$U(z) = \frac{u_*}{\kappa} \ln \frac{z}{z_0} \quad (5)$$

式中: $U(z)$ 为高度 z 处的10 min平均风速; z_0 为地面粗糙度。

根据式(5)不难理解,风剖面的确定仅需要2个独立参数,本文采用剪切波速 u_* 与地面粗糙度 z_0 这一组合作为风剖面基本参数.在物理意义上加以考察,这种选择是合理的:剪切波速 u_* 本质上反映了涡动黏性力在大气边界层中对平均风速的滞变效应,它建立起风的平均大尺度特性与高频细部结构之间联系的桥梁^[6];地面粗糙度 z_0 则将边界条件引起的底部涡旋以特征长度的方式计入风剖面,它实质上是边界条件产生的阻力对平均风速的影响^[17].这样,加上分界波数 k_c ,本文所研究的风速模型有3个基本随机变量。

2 参数识别

2.1 厦门风场观测台阵

本研究小组于2014年在厦门大嶼岛(北纬 $24^{\circ}33'9''$,东经 $118^{\circ}19'47''$)建立了一座强风观测台阵.这一台阵共有4座观测塔P1、P2、P3、P4,它们的空间布置如图2所示.在P1塔与P4塔的10 m与20 m处安装有三维超声风速仪,在P2塔与P3塔的10 m、20 m、30 m及40 m处安装有三维超声风速仪,采样频率均为10 Hz.观测台阵周围以农业用地与灌木为主,在其四周100~200 m远处分布着多个建筑密集区,多为10 m以下民房,总体上该地区的地貌与《建筑结构荷载规范》^[18]的B类地貌或C类

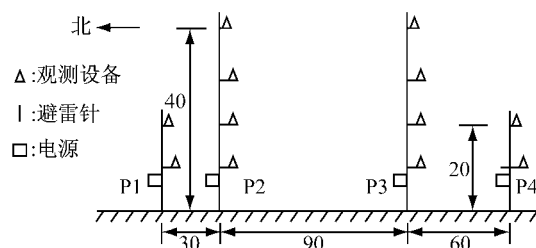


图2 风场观测台阵的布置(单位:m)

Fig. 2 Lay-out of the strong wind observation system (unit: m)

地貌相当.在本文的分析中,采用了2015年6月28日至2015年8月19日之间的数据,经过检查,所采用数据都处于良态风环境.经过对比发现:由其他各塔的数据得出的结论与P3塔相一致,因此这里只讨论P3塔的结果。

2.2 参数识别方法与结果

本文采用最小二乘法进行参数识别.这一方法实质上是在合适的距离定义下寻找实测样本的最佳逼近.考虑实测样本 y 与模型 $y(\Theta)$ 的二范数误差

$$J(\Theta) = \|y - y(\Theta)\|_2 \quad (6)$$

式中: y 是观测值; $y(\Theta)$ 是物理模型.参数识别结果 Θ_{id} 应使式(6)取得最小值.如果将样本 y 与模型 $y(\Theta)$ 分别取为实测脉动风速的Fourier幅值谱与双线性Fourier幅值谱(式(2)),就能够识别出其基本参数剪切波速 u_* 与分界波数 k_c ;同理,可以进行对数律风剖面的识别。

应当指出,根据脉动风速Fourier幅值谱(式(2))与风剖面(式(5))都可以识别得到剪切波速.如前所述,在Fourier谱中,剪切波速反映了考察位置处涡动黏性引起的主流剪切;而对数律中剪切波速表征了竖向剪切的平均.因此,本文先识别了各高度处Fourier幅值谱中的剪切波速;然后取其平均值作为风剖面的剪切波速进一步识别地面粗糙度.图3和图4是同一10 min间隔内的典型脉动风速Fourier幅值谱及风剖面的识别结果.可以看到,各高度处的剪切波速十分接近,说明这一识别方法是合理的。

将所有识别结果作统计分析, u_* 、 z_0 及 k_c 的结果分别见图5、图6与图7.与以前研究结果相比^[19],关于 u_* 识别结果的统计分布没有本质差别.但大量 z_0 集中在10 m左右,显然不符合实际地面粗糙度分布范围;此外3个高度 k_c 的识别结果中出现了明显的双峰特征,这表明分界波数识别结果中可能存在特有的数据结构。

为初步检验由上述识别结果得到的概率分布对风速时程基本概率信息的影响,可以考察它们所导出的功率谱密度函数并与经典谱对比.这里取Kaimal谱作为参照

$$\frac{nS(z, n)}{u_*^2} = \frac{200f}{(1 + 50f)^{5/3}} \quad (7)$$

式中: $f = nz/U$; $S(z, n)$ 为脉动风速的功率谱.注意到功率谱是样本Fourier谱的平均

$$S(n) = 2E[|F(\Theta, n)|^2] \quad (8)$$

采用MATLAB中概率分布拟合的核密度估计方法,给出各高度分界波数的概率密度函数,如图7

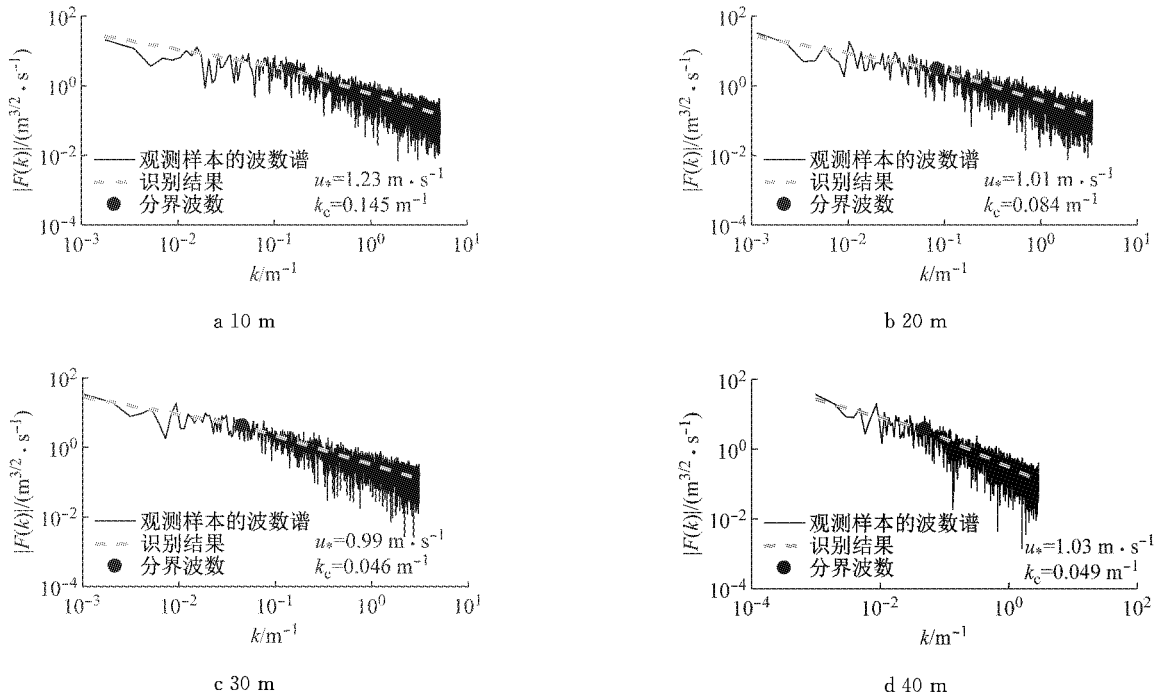


图 3 典型波数谱与识别结果

Fig. 3 Typical wave number spectrum and identification

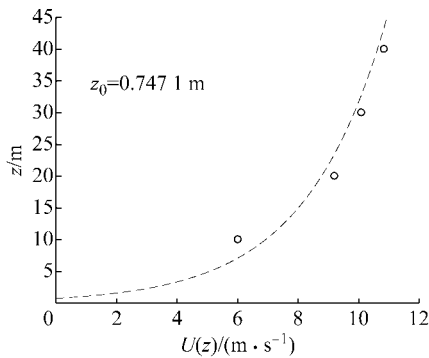


图 4 典型风剖面与识别结果

Fig. 4 Typical mean wind profile and identification

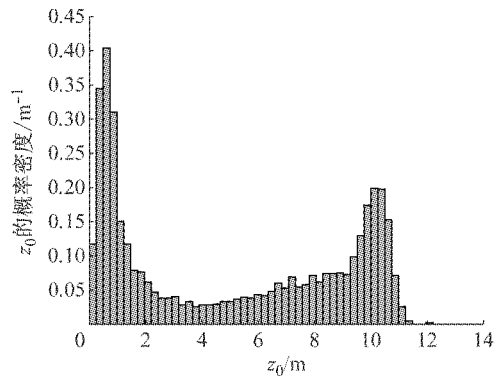


图 6 地面粗糙度统计结果

Fig. 6 Statistics of roughness length

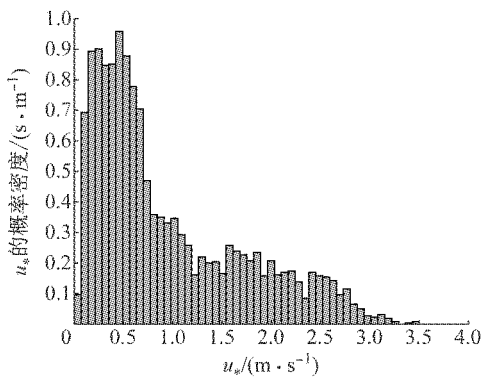


图 5 剪切波速统计结果

Fig. 5 Statistics of shear velocity

所示. 在此基础上, 结合式(8)计算了由随机 Fourier 谱模型导出的量纲一功率谱, 如图8所示. 显然, 在

高频段模型结果与 Kaimal 谱基本一致, 但在中、低频段很大范围内, 由分界波数直接识别结果导出的功率谱密度函数显著低于 Kaimal 谱.

为深入研究上述问题, 在图 9 的各子图中作出不同高度分界波数的关系, 其中各点的横、纵坐标为相应时刻不同高度分界波数. 显然, 任意两个高度处分界波数组成的点集都泾渭分明地分为两簇. 这一特征及其原因可以通过聚类方法进行分析.

3 分界波数的聚类分析与建模

3.1 聚类分析的基本原理

已经指出, 聚类分析能够利用样本属性的相似性对样本集合进行分类. 本文采用欧几里得距离作

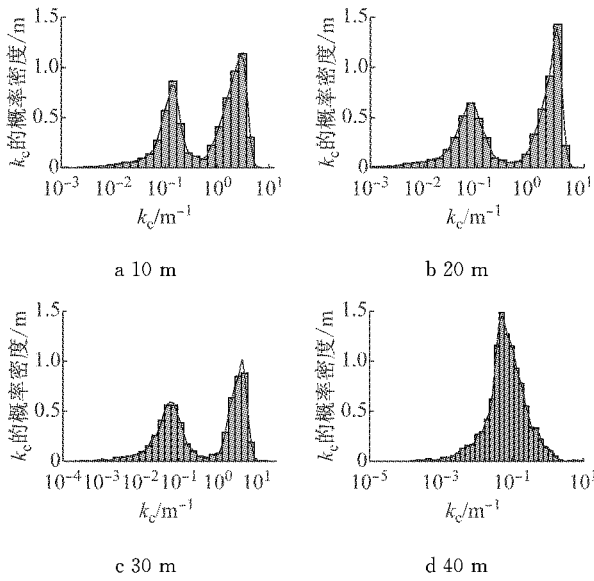


图 7 分界波数统计结果

Fig. 7 Statistics of cut-off wave number

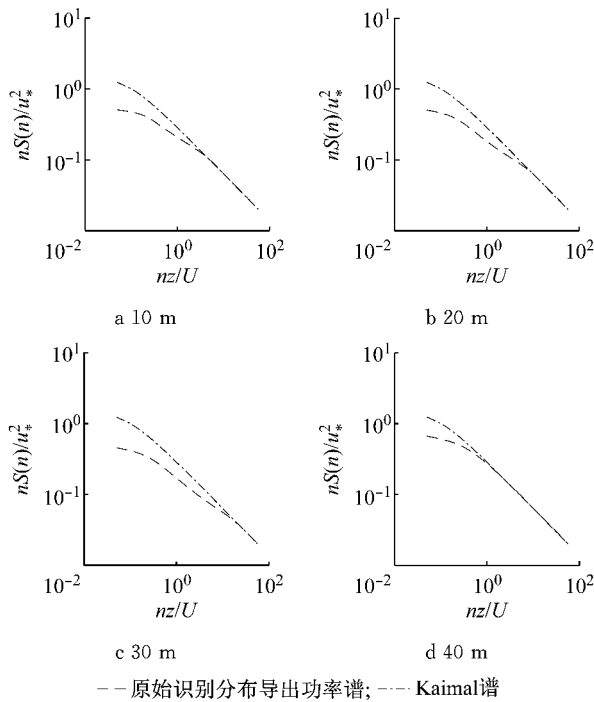


图 8 由原始识别结果拟合分布导出的功率谱

Fig. 8 Power spectrum density derived from statistical distribution of original identification

为样本相似性的描述,即:将 n 维点集 $X = \{x_i, i = 1, 2, \dots, n\}$ 中的点分成 N 类 $\{C_l, l = 1, 2, \dots, N\}$,使得总偏差最小,总偏差的定义为

$$D(C) = \sum_{l=1}^N \sum_{x \in C_l} \|x - \mu_l\|^2 \quad (9)$$

式中: μ_l 是第 l 类的中心点; $\|\cdot\|$ 是欧几里得距离.

一般可以使用经典的递归算法实现上述目标:

(1) 首先适当地选取 N 个中心点,作为各类的

特征点,将集合 X 中的点按照距所选择的中心点的远近将其归类,接下来重复(2)与(3)的操作,直至分类最终稳定;

(2) 按上一次分类的结果重新计算 N 个中心点,即每一类所有点的平均值;

(3) 按新的中心点集进行分类.

根据对图 9 的分析,本文分类数 N 取为 2 即可.

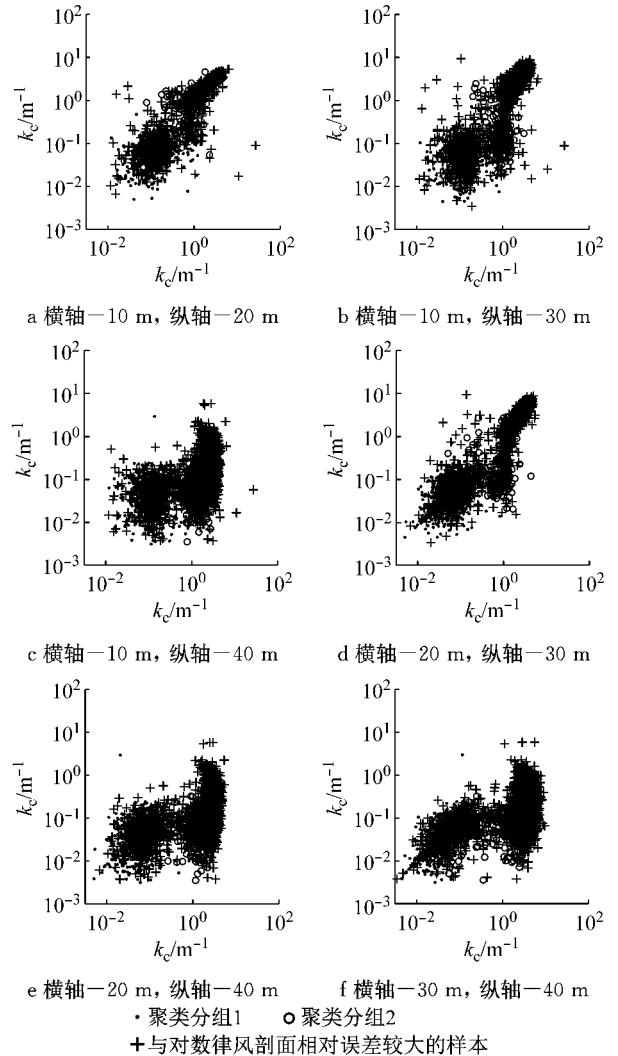


图 9 分界波数识别结果

Fig. 9 Cluster results of cut-off wave number

3.2 分界波数的聚类结果

将同一时间段内、4 个高度识别出的分界波数看作 R^4 空间(每一维度代表某一高度的分界波数值)的一个点,则所有记录得到的 4 维分界波数数组全体构成了待聚类的点集.本文利用 MATLAB 中 k -means 函数进行聚类,聚类结果如图 9 所示.显然,通过聚类分析,上节所述的分类特征得以细致地展现.需要说明的是,图 9 中“+”代表实测风剖面与

对数律风剖面相对误差大于 5.5% 的样本,而聚类分析只考虑相对误差小于 5.5% 的样本.

3.3 聚类现象原因分析

可以推断,图 9 中“o”形点集受到了外界因素的扰动.一般,风场中扰动源常有固定的空间位置,故可以在风向图中探究样本聚类的原因.

在极坐标图 10 中作出各高度 10 min 平均风速 U_{10} 与风向的关系,并用相应的记号表现上述聚类结果,对图 10 的考察可以得出以下规律:

(1) 风剖面受显著干扰的样本几乎完全集中在 0° (正北)方向以及附近区域;

(2) 与其他方向比较,平均风速在 0° (正北)方向有明显的凹进,说明它在此处被显著削弱;

(3) “o”形点集以带状形式分布在上述大相对误差样本两侧,它们一起占据了与 0° (正北)方向左右约 60° 的范围;“·”形点集则散布在上述范围外.

综合上述结果,可以推测:当风从正南方吹来时,受到某种因素的扰动,风速被削弱,且风速受到的扰动随着风向与正北方的夹角增大而减小.考察风场观测台阵设备设施,注意到各高度处的三维超声风速仪都安装在由观测塔朝正北方向伸出约 1 m 左右的构件上,而且观测塔直径也在 1 m 左右,则有理由相信,观测塔是扰动实测记录的主要因素.

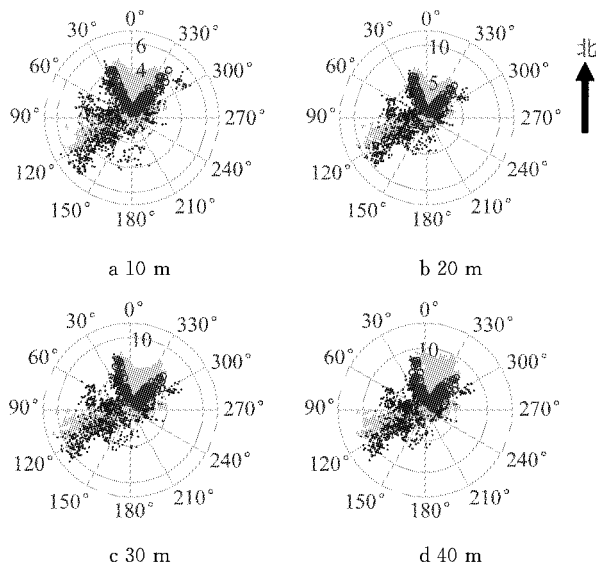


图 10 分界波数聚类结果与风向的关系(图中半径大小代表 10 m 高、10 min 平均风速,单位: $m \cdot s^{-1}$)

Fig. 10 Relation of cluster results to wind direction (The magnitude of radius designate 10 min mean wind speed at a height of 10 m, unit: $m \cdot s^{-1}$)

在已有的风场数据分析文献中,鲜有考虑观测

设备设施对实测记录影响的工作.本文工作说明,为了获得正确合理的结论,应当重视观测系统自身带来的误差.

3.4 基于聚类分析的分界波数建模

前文分析指出,将分界波数处理为随机变量是一种更合理的建模方式.为了正确地统计其概率分布,应当将上文讨论的外界扰动排除在外.据此,本文将风向在正北方向附近一定方位角内及风剖面与对数律相对误差大于 5.5% 的样本记录排除.由此得到的地面粗糙度的直方图见图 11,以及各高度处分界波数的对数直方图见图 12.可见,图 6 的地面粗糙度异常分布及图 7 分界波数的双峰分布不再存在.

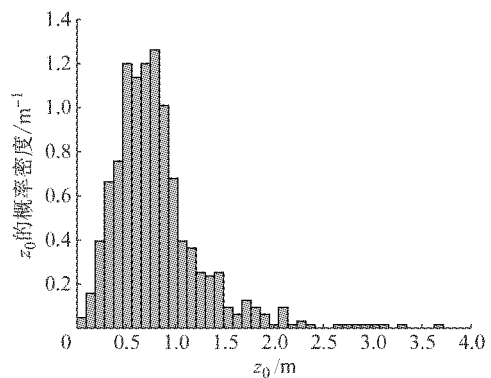


图 11 数据筛选后的地面粗糙度统计直方图
Fig. 11 Histogram of roughness length after data sieving

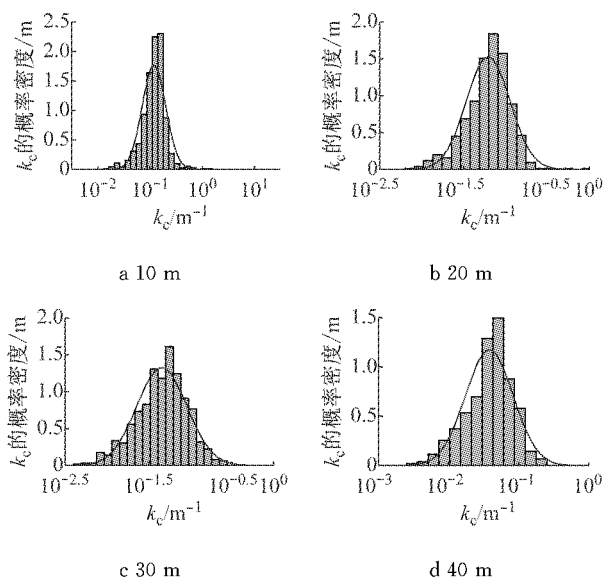


图 12 分界波数直方图与建模

Fig. 12 Histogram and modeling of cut-off wave number

进一步考察样本筛选是否能够改善功率谱密度的模拟.采用对数正态分布对分界波数进行建模,建模结果如表 1 所示.分析可见, k_c 对数均值 $\mu_{\ln k_c}$ 与对

数标准差 $\sigma_{\ln k_c}$ 与高度有明显的线性关系(见图 13 与图 14).

$$\mu_{\ln k_c} = -0.016 3z - 0.809 8$$

$$\sigma_{\ln k_c} = 0.003 6z + 0.186 2$$

式中:高度 z 单位为 m; k_c 单位为 m^{-1} .

表 1 分界波数建模结果

z/m	$\mu_{\ln k_c}$	$\sigma_{\ln k_c}$
10	-0.916 2	0.222 2
20	-1.205 4	0.257 8
30	-1.333 9	0.290 4
40	-1.418 1	0.329 7

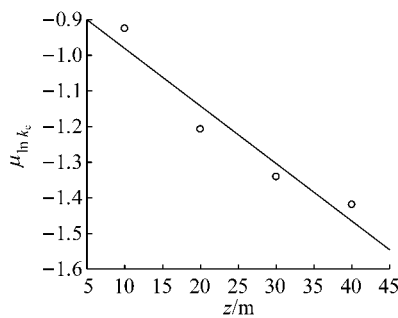


图 13 分界波数对数均值与高度的关系

Fig. 13 Logarithmic mean value of cut-off wave number versus height

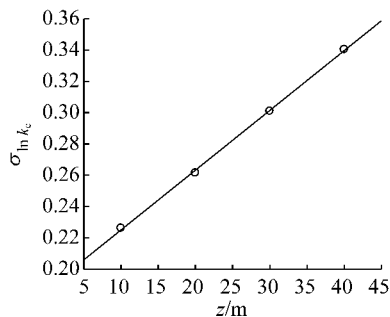


图 14 分界波数对数标准差与高度的关系

Fig. 14 Logarithmic standard deviation of cut-off wave number versus height

按 2.2 节方法计算由随机 Fourier 谱模型导出的功率谱,并与 Kaimal 谱对比,如图 15 所示.显然,排除异常数据后,模型功率谱与 Kaimal 谱在一般工程所关心的频率范围内几乎一致,证明聚类分析与样本筛选的有效性.同时,上述分析进一步验证了随机 Fourier 谱模型对脉动风速细部特征进行合理刻画的能力.

4 结论

本文研究了脉动风速随机 Fourier 谱模型基本

随机变量及其分布,得到以下结论:

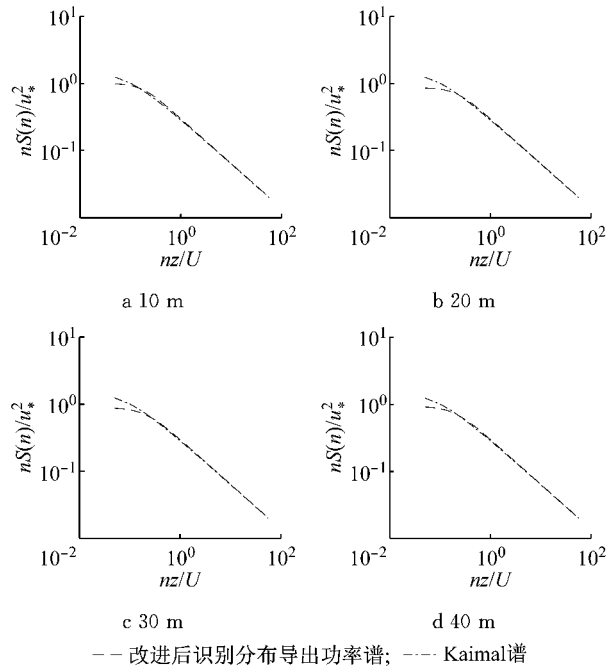


图 15 数据筛选后的参数统计分布导出的功率谱

Fig. 15 Power spectrum density derived from statistical distribution after data sieving

(1) 脉动风速随机 Fourier 谱模型的基本随机变量包括剪切波速、地面粗糙度、分界波数以及零点演化时间.利用厦门地区实测风速记录,识别了剪切波速、地面粗糙度、分界波数.

(2) 初始识别结果中地面粗糙度分布异常;分界波数呈现双峰特征.初始识别结果的统计分布导出的功率谱密度函数明显偏离 Kaimal 谱.

(3) 研究发现,上述参数识别结果的异常源自风速样本的聚类特征,且与风向紧密联系.这一现象与观测装置对风场的扰动有关,因此在分析实测风场记录时,应重视风场设备设施对观测效果的影响.

(4) 依据聚类分析结果,合理地选择样本集合,进行了分界波数统计建模.由此导出的模型功率谱与 Kaimal 谱对比良好,这也证明了聚类分析的有效性.

参考文献:

[1] JAIN A K. Data clustering: 50 years beyond K-means[J]. Pattern Recognition Letters, 2010, 31(8): 651.

[2] 方开泰,潘恩沛. 聚类分析[M]. 北京:地质出版社,1982. FANG Kaitai, PAN Enpei. Cluster analysis [M]. Beijing: Geological Publishing House, 1982.

[3] 苍宏宇,谭宗颖. 聚类搜索引擎发展现状研究[J]. 图书情报工作, 2009, 53(2):125.