

基于长期连续测波资料的波浪平稳性分析

武雨晴^{1,2}, 陶爱峰^{1,2}, 文铖², 李硕², 曹广随²

(1. 河海大学 海岸灾害及防护教育部重点实验室, 江苏 南京 210098; 2. 河海大学
港口海岸与近海工程学院, 江苏 南京 210098)

摘要: 海浪的平稳性是研究其他海浪问题的前提条件, 然而现实中的海浪并不是严格平稳的, 若测波时间太长, 其平稳性会受到影响。因此在分析海浪长期连续实测数据时, 要保证海浪记录采用的时间长度满足平稳性的要求。利用韩国周边海域9个波浪测站, 每个测站测波时长从1~12个月不等的长期连续资料分析波浪样本的平稳性。对每个测站每个月的波浪资料按不同的时间数据长度进行分段, 用图检验和统计检验的方法计算了不同时间长度下的每一段波浪样本的平稳性, 并对比了检验效果。结果表明: 从时序图、自相关图中不能发现不同时间长度情况下的平稳性差异, 且从图中看不出数据是否存在随时间变化的单调趋势; 而统计检验中的游程检验不适合用于波浪样本检验; 逆序检验法的结果表明当韩国各测站每月的测波资料按20min分段时, 有6%~16%的序列不平稳, 但相对于采用其他时间长度分段的情况, 其不平稳序列个数较少。

关键词: 实测数据; 平稳性; 图检验; 统计检验

对于随机波浪的研究, 通常采用谱方法及统计分析方法确定波高、周期等特征值, 然而这两种方法都是建立在平稳数据的基础上^[1,2]; 且任何波浪理论研究或工程实际应用都离不开波浪特征值的确定, 由此可见平稳性是研究随机波浪理论的基础条件。所谓平稳性是指随机序列的统计特征不随时间的推移而发生变化^[3]。理论上讲, 时间越长, 统计特征越稳定, 然而海浪并不是严格平稳的, 记录时间太长可能会影响其平稳性, 因此对于长期连续测波资料的分析, 首先要保证其平稳性^[4]。

韩国海洋科学与技术研究所(KIOST, 原KORDI)在韩国周边海域陆续设立了多个波浪连续观测站, 积累了大量的波浪资料, 部分学者^[5,6]对这些连续资料采用30min分段来获取其有效波高等波浪特征值, 但并没有相关理论依据, 为此本文将利用韩国周边测站2009年的长期连续观测资料对该区域的波浪平稳性进行分析。

1 随机过程的平稳性

平稳性分为严平稳性和宽平稳性也称为弱平稳性。严平稳性是指随机序列的所有统计特征都不随时间的变化而变化, 其中统计特征是由随机序列的联合分布函数所决定的, 而在实际中要想获得随机序列的所有联合分布函数几乎是不可能的, 且在通常情况下只需关注随机序列的数字特征即可, 这是由序列的低阶矩(一阶、二阶矩)所决定的^[3]。因此当随机序列的一阶矩、二阶矩不随时间发生变化时称该序列为宽平稳序列, 目前所提到的平稳性通常指的都是宽平稳性。若随机序列 $\{X_t, t \in N\}$ 满足条件1)~3), 则认为该随机时间序列是平稳的, 而对应的随机过程为平稳随机过程^[7]。

1) 对于任何的 $t \in N$, 随机序列的期望是与时间 t 无关的常数:

$$E(X_t) = \mu = \text{常数} \quad (1)$$

2) 对于任何的 $t \in N$, 随机序列的方差是与时间 t 无关的常数:

$$\text{Var}(X_t) = s^2 = \text{常数} \quad (2)$$

3) 对于任何的 $t, k \in N$, 随机序列在 $t, t+k$ 时刻的协方差是只与时间间隔 k 有关, 与时间 t 无关的常数:

$$\text{Cov}(X_t, X_{t+k}) = gk \quad (3)$$

基金项目: 国家自然科学基金资助项目(41106001, 51137002); 教育部留学回国人员科研启动基金资助项目(教外司留[2012]1707)

2 数据介绍

本文所用数据来源于韩国周边海域陆 9 个波浪测站 2009 年的实测数据，测站位置见图 1。测量仪器采用水下波压计 (WTG-128M)，埋设在距海底 50 cm 处，设置采样频率为 2Hz。每个测站数据时间长度为 1~12 个月不等。由于原始波压数据是每 0.5 s 一个信号的连续性资料，平均每组数据达到 2.5×10^7 ，为了方便计算，先将数据按 30 d 为一个单位进行分组，之后再按平稳性进行进一步分组。由于原始信号为波压信号，而波压信号转为波面信号存在一定的误差，且本文只是对韩国实测数据的平稳性进行分析，并不涉及波浪的相关理论研究，因此不需将波压信号转为波面信号。

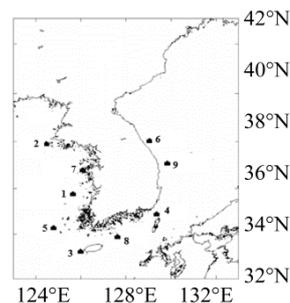


图 1 韩国周边海域 9 个测站位置

3 平稳性检验

对于随机序列的平稳性检验分为两种检验方法，一种是根据时序图和自相关图显示特征进行判断的图检验法；另一种是通过构造检验统计量进行假设检验的方法，包括逆序法、游程检验法^[8]。不同检验方法各有优缺点，要根据实际数据的具体特征选取合适的方法进行检验^[9]。本文以韩国 Anmado 测站第 1 个月波压数据为例，按不同时间长度分别对该月数据进行进一步分组，并运用不同方法进行平稳性检验以确定韩国周边海域波浪实测数据较为合适的样本长度。

3.1 时序图检验

时序图是指将实测数据按照时间先后顺序连成曲线，绘制成以时间为横坐标，实测数值为纵坐标的二维平面图形。根据平稳随机序列均值、方差为常数的性质，平稳序列的时序图应表现出所有数据围绕一常数上下波动且波动范围有界的特点^[8]。将 Anmado 测站第 1 个月数据分别按 5~30min 分组，其第一段序列的时序图如图 2 所示。从图中看出，所有序列的时序图都符合平稳序列的特点且不同分组情况下的时序图并无差异。因此，仅从时序图特征不能准确判断不同情况下随机序列是否平稳。

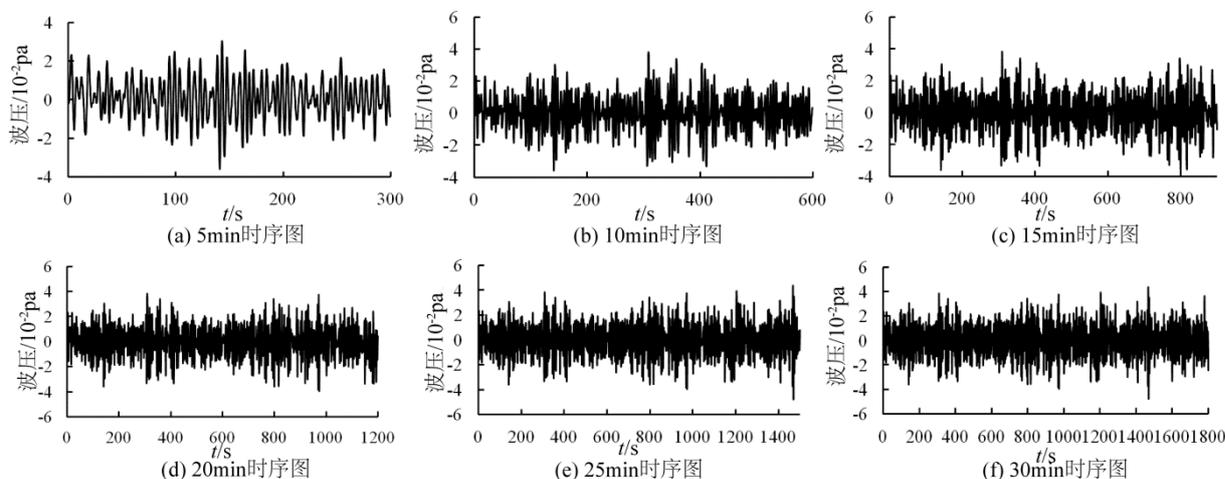


图 2 Anmado 测站第 1 个月数据不同分段长度下的时序

3.2 自相关图检验

自相关图是指以时间间隔 k 为横坐标，样本自相关系数为纵坐标的二维平面图。对于随机序列 $\{X_t, t \in N\}$ ，样本自相关系数的计算式为^[10]：

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^N (X_t - \bar{X})^2} \quad (4)$$

式中: N 为数据总长度; t 为时间; k 为时间间隔。平稳随机序列具有短期相关性的特点, 在自相关图中表现为随着时间间隔 k 的增大, 自相关系数迅速衰减为 0 的特点^[8]。将 Anmado 测站第 1 个月数据分别按 5~30min 分组, 其第一段序列前 300s 的自相关图分别如图 3 所示。从图中看出, 各序列随着时间间隔的增大, 自相关系数衰减为 0 的程度无明显差异。因此, 仅从自相关图特征不能准确判断不同情况下随机序列是否平稳。

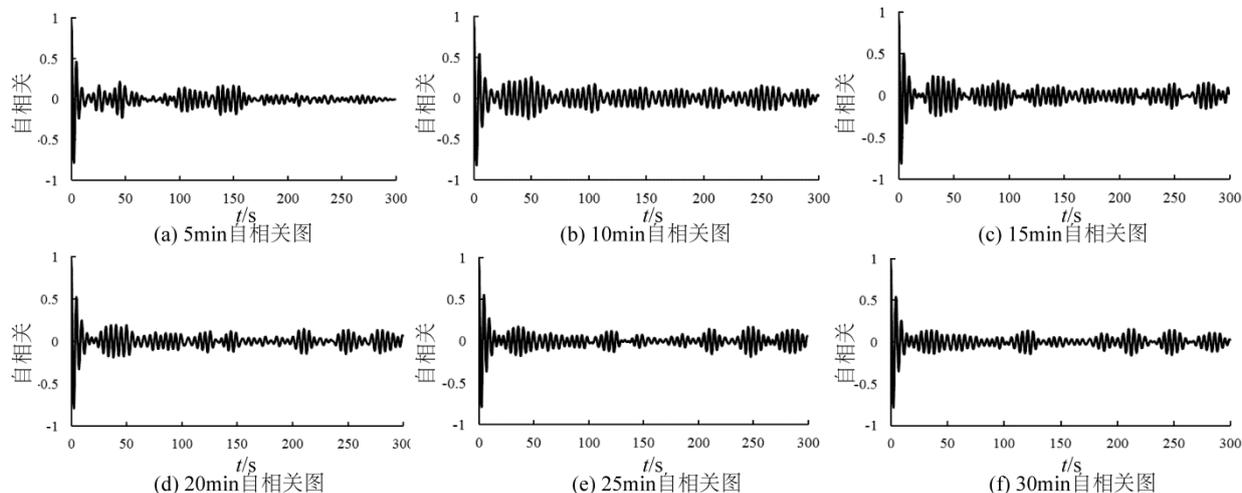


图 3 Anmado 测站第 1 个月数据不同分段长度下的自相关图

3.3 游程检验

设随机序列 $\{X_t, t \in N\}$ 的均值为 \bar{X} , 将序列中比 \bar{X} 小的值记为“—”号, 大的值记为“+”号, 连续相同的一段记号记为一个游程, 用 r 表示。若游程数 r 太小, 表明观测值连续大于或低于均值, 序列中可能存在某种趋势; 而游程数 r 太大, 说明该序列中也可能存在某种此消彼长的必然趋势; 因此对于平稳随机序列来说, 各观测值应在均值附近上下波动, 大于或小于均值的个数应该适当, 对应的游程数 r 值也应适当^[10]。

对于随机序列, 当样本长度较少时, 游程数 r 服从 r 分布:

$$E(r) = \frac{2N_1N_2}{N} + 1 \tag{5}$$

$$D(r) = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N - 1)} \tag{6}$$

当样本数较多时, 构造新的统计量 Z , Z 服从标准正态分布:

$$Z = \frac{r - E(r)}{\sqrt{D(r)}} \sim N(0,1) \tag{7}$$

式中: N_1, N_2 分别为“—”、“+”的个数, 在给定显著性水平 α 下, 当 N_1 和 N_2 均小于 15 时, 游程数 r 服从 r 分布, 平稳随机序列的游程数应在一定的范围内, 该范围可由 r 分布表查得; 而当 N_1 和 N_2 均大于 15, 在给定显著性水平 $\alpha = 0.05$ 时, 平稳随机序列统计量 $|Z| < 1.96$ 。

将 Anmado 测站第 1 个月数据分别按 5~30min 分成若干段序列即样本, 计算每段序列的平均值, N_1 和 N_2 的值以及游程总数 r , 用上述方法计算各序列是否平稳, 其计算结果见表 1。

从表 1 发现, 在不同分段长度情况下该月所有序列都是不平稳序列, 而在实际海浪中波面高度的平均值应为 0, 即不随时间而发生变化, 符合平稳性要求。因此计算结果与实际不符, 且经其他学者验证^[11], 游程检验并不适用于周期性数据的平稳性检验。

表 1 游程检验 Anmado 测站第一个月数据在不同分段长度下的不平稳序列比例

分段长度/min	不平稳序列比例/(%)	分段长度/min	不平稳序列比例/(%)
5	100	20	100
10	100	25	100
15	100	30	100

3.4 逆序法检验

设随机序列 $\{X_t, t \in N\}$ ，当 $t > s$ 时，如果满足 $X_t > X_s$ ，则称为 1 个逆序。设 A_t 为每个实测数据的逆序数，则 A 为逆序总数：

$$A = \sum_{t=1}^N A_t \tag{8}$$

将随机序列分为互不相关的 M 段，求出每一段的均值或方差，并构成新序列 $\{y_t, t=1, \dots, M\}$ ，该序列中的变量为 M 个均值或方差，计算均值或方差序列的逆序总数 A ，构造统计量 Z 并进行统计检验^[12]：

$$Z = \frac{A + \frac{1}{2} - E(A)}{\sqrt{D(A)}} \sim N(0,1) \tag{9}$$

$$E(A) = \frac{1}{4} M(M-1) \tag{10}$$

$$D(A) = \frac{M(2M^2 + 3M - 5)}{72} \tag{11}$$

式中： $E(A)$ ， $D(A)$ 分别为逆序总数的期望与方差。

逆序检验法是为了检验随机序列的均值或方差是否存在随时间变化的趋势，若序列平稳，则 M 个均值或方差应随机分布，逆序总数 A 为大小适中的值，在给定显著性水平 $\alpha=0.05$ 下，统计量 $|Z| < 1.96$ ；若序列表现出随时间变大或变小的趋势，则均值或方差的逆序数偏大或者偏小^[3]。

将 Anmado 测站第一个月数据分别按 5~30min 分成若干段序列，再将每段序列分为 30 段，即 $M=30$ ，计算每段的平均值、方差，以及逆序总数 A ，用上述方法计算各序列是否平稳，其计算结果见表 2。

表 2 均值和方差逆序检验 Anmado 测站第 1 个月数据在不同分段长度下的不平稳序列比例

分段长度/min	不平稳序列比例/(%)	分段长度/min	不平稳序列比例/(%)
5	0.09/15.13	20	0/11.53
10	0/11.90	25	0/12.96
15	0/10.28	30	0/13.61

注：“/”前面数据为均值逆序，后面数据为方差逆序。

从表 2 中看出，当 Anmado 测站第 1 个月数据按任意长度分段时，其每个均值序列都符合平稳性要求，这与实际海浪中波面高度均值为 0，不随时间变化的理论一致；当 Anmado 测站第 1 个月数据按 15min 分段时，不平稳序列占总序列的比例较少，随着分段长度的增大，不平稳序列所占比例也随之增大。波浪样本中波数不宜少于 100^[4]，而该月数据按 15min 分段时，有 19.93% 的序列单波数小于 100，而按 20min 分段时，波数小于 100 的序列较少，平稳序列相对其他分段情况下所占的比例较大。因此，综合考虑，Anmado 测站第 1 个月数据按 20min 数据长度进行进一步分组时，该月数据平稳性及单波数都相对较好。

运用该方法计算其他测站各月份数据在不同分段长度下随机序列的平稳性，由于海浪波面高度的平均值为 0，因此只需用逆序法检验方差平稳性即可。以 Anmado 测站第 4 个月数据为例，逆序检验方差平稳性的计算结果如表 3。

表 3 方差逆序检验 Anmado 测站第 4 个月数据在不同分段长度下的不平稳序列比例

分段长度/min	不平稳序列比例/(%)	分段长度/min	不平稳序列比例/(%)
5	15.28	20	7.96
10	9.49	25	9.72
15	9.86	30	10.00

由表3得到该月数据按20min分段时,平稳序列比例较多,所有序列的单波数都大于100,因此该月数据按20min分段较合适,这与该测站第1个月的平稳性分段长度不同。由其他测站各月份的计算结果发现,每个测站每个月按平稳性划分的序列长度各不相同,综合来看,当按20min数据长度分组时,各测站有6%~16%的序列不平稳,而相对于其他分段长度的情况,该平稳性比例相对较好,且单波数小于100的序列较少。因此建议韩国周边海域各测站数据按20min进行分组以各测站第1个月数据为例,按20min数据长度分段时各测站该月不平稳序列比例如表4所示。

表4 方差逆序检验各测站第一个月数据在20min分段长度下的不平稳序列比例

测站	测站名称	不平稳序列比例/(%)	测站	测站名称	不平稳序列比例/(%)
1	Anmado	11.53	6	Kangrung	7.78
2	Baerueongdo	15.05	7	Malipo	15.32
3	Biyangdo	8.70	8	Sorido	8.75
4	Hongdo	9.31	9	Wandolcho	7.64
5	Ilhyangcho	7.18			

4 结 语

本文利用韩国周边海域多个测站2009年的波浪长期连续观测资料,运用时序图检验、自相关图检验、游程检验、逆序法检验的方法对随机波浪序列在不同分段长度下的平稳性进行检验,并将不同方法的检验结果进行对比。结果表明,时序图检验、自相关图检验虽简单明了,但不够精确,主观性较强,从图中不能明显判断各参数的平稳性差异;而统计检验的结果较为准确,但游程检验并不适用于周期性数据的检验;逆序检验对于该序列的平稳性检验较为合适。综合对韩国周边海域各测站各月份的平稳性逆序法检验结果,最终认为按20min数据长度分段时序列的平稳性较好。

参考文献

- 1 杨秉正. 随机风浪资料的数据分析和检验[J]. 内蒙古水利科技, 1988: 14-17.
- 2 杨秉正. 随机风浪数据的频谱分析[J]. 内蒙古水利科技, 1989, 2(002): 13-17.
- 3 史代敏, 谢小燕. 应用时间序列分析[M]. 北京:高等教育出版社, 2011: 13-177.
- 4 俞聿修, 柳淑学. 随机波浪及其工程应用[M]. 大连:大连理工大学出版社, 2000: 187.
- 5 JEONG W M, OH S H, RYU K H, et al. Multi-station observation of the high swell along the east coast of Korea[C]//Proc. 5th Int Conf Asian Pac Coasts, Singapore, 2009.
- 6 OH S H, JEONG W M. Characteristics of high waves observed at multiple stations along the east coast of Korea[J]. Natural Hazards and Earth System Science, 2013, 13(12): 3503-3514.
- 7 范涛涛, 寇艳廷, 刘晨, 等. 时间序列分析中数据的平稳性判定研究[J]. 现代电子技术, 2013, 36(4): 66-68.
- 8 王燕. 应用时间序列分析[M]. 北京:中国人民大学出版社, 2005: 23-25.
- 9 吕光明. 对数据平稳性检验方法的比较研究[J]. 财经问题研究, 2004(6): 77-80.
- 10 肖枝洪. 时间序列分析与SAS应用[M]. 武汉:武汉大学出版社, 2012: 16-27.
- 11 李庆雷, 马桶, 付遵涛. 时间序列非平稳检测方法的对比分析[J]. 北京大学学报:自然科学版, 2013(2): 252-260.
- 12 刘罗曼. 时间序列平稳性检验[J]. 沈阳师范大学学报:自然科学版, 2010, 28(3): 357-359.