

# 基于 M-K 方法及成分提取的海洋表层热状分析

韩心娅 沈 翀 沈婷婷 程宗毛

(杭州电子科技大学理学院, 浙江 杭州 310018)

**摘要:** 本文针对东海一点的 9 年海表温度遥感数据集的序列 (2003-2011) 进行处理, 首先通过 SPSS 初步检验出所处理的温度数据满足正态性。然后通过 Mann-Kendall 秩次相关检验研究了数据点温度的趋势性和突变性。通过周期存在性检验取定四个周期。用傅里叶函数描述周期成份, 各自系数分别通过参数估计得到。对比测试数据的周期项和原始数据图表得结果有效。该随机序列的自相关是拖尾的, 而偏相关是一阶截尾的, 故可建立 AR (3) 模型, 分析得序列是白噪声序列。

**关键词:** 海表温度; Mann-Kendall 方法; 趋势性; 突变性; 周期成分

**1. 引言:** 时间序列分析是概率统计学科中的一个分支, 随着基本理论的更加深入和完善, 以及统计软件技术的发展, 使其在气象水文等领域有了更加广泛的应用。海表温度不仅是描述海洋表层热状况的重要指标, 其异常情况还是海洋影响大气环流、气候变化的主要因子。所以海表温度一直都是诸多学者观察、研究和预测的重要研究对象。趋势性、突变性和周期性是时间序列数据的基本特征, Mann-Kendall 方法作为一种非参数统计检验方法, 其优点在于样本不需遵从一定的分布, 也不会受少数异常值的干扰, 计算相对方便。近几年被广泛用于气温、降水量变化、植被变化等自然界变化的趋势分析中, 考虑到海表温度的变化规律也有相似的性质。因此, 本文选择通过处理海表温度使其可以显示出正态性, 而后引入该 M-K 方法测试数据在一定时间段内是否异常, 在此基础上再进一步提取海表温度集的周期成分, 确立时间序列模型。

## 2. 数据预处理

本文考虑建立合理的分析模型提取海洋表层温度数据的趋势性、周期性和突变性特征, 通过误差分析比较不同提取方法的优缺点, 为预测海表温度未来的变化提供更加有效的依据。此外, 将温度数据集进行合理的区间划分, 使得每个区间内的统计数据复合最优的正态分布统计, 并给出正态分布的检验方法、统计区间的上下节点及统计分布模型的相关参数。

本文采用了东海海表上 2003 年至 2011 年一研究点的温度遥感数据集, 从 1 开始编号, 共计数据 3287 个。计一年为 365 天, 这组数据记录了 9 年里每一天的温度, 但是由于 2004 年和 2008 年是闰年, 为减小不同年份的对比差异, 将这两年 2 月 29 日的温度数据剔除。推算可得这两个时间点的温度数据对应的序号分别为 425 和 1886, 剔除后数据长度变为 3285。

几乎所有的科研数据都必须满足正态性才能进行分析, 因此要对数据进行正态性检验, 以保证后续分析的可靠性。本文利用 SPSS 对每个点每一年的温度数据进行了直方图和正态曲线绘制, 具体将东海 2003 年的情况呈现如下:

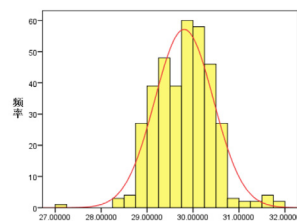


图 1 东海 2003 年数据正态分布图

东海	
N	365
正态参数 a,b	均值 29.8065716
	标准差 .63733542
最极端差别	绝对值 .055
	正 .055
	负 -.037
Kolmogorov-Smirnov Z	1.043
渐近显著性(双侧)	.227

图 2 东海 2003 年数据正态性检验

观察图 1 可得数据呈正态分布。图 2 给出了 K-S 检验的具体结果, 由于概率 p 值为 0.227, 大于 0.05, 故不能拒绝 K-S 检验的原假设, 即数据满足正态性。这为接下来的探讨提供了一定的支持和帮助。

## 3. 海表温度基本特征分析

**3.1 Mann-Kendall 方法简介。** Mann-Kendall 方法由于最初由 H.B. Mann 和 M.G. Kendall 提出原理并发展了该方法。当时这一方法仅用于检测序列的变化趋势, 后来经他人进一步完善和改进, 才形成目前的计算格式。该方法既可以检测序列的变化趋势, 也可以进行突变点检验。

### 3.2 Mann-Kendall 方法应用

**3.2.1 研究点海表温度的日变化趋势与突变情况。**在对东海表温度的数据的趋势性和突变性信息进行提取时, 本文主要采用 M-K 方法, 通过 matlab 软件编写程序, 我们得到下述结果:

#### ①趋势变化:

表 1 海表温度日变化数据分析表

n	Upward trend detected		Sens Nonparametric Estimator		
	Mean Value	Z statistic	斜率估计	置信下限	置信上限
3285	29.6398	4.381	5.7971e-005	3.1194e-005	8.3631e-005

#### ②突变情况:

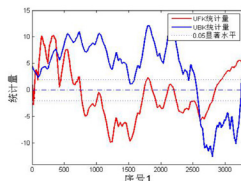


图 3 海表日温度突变情况图

**③结果分析:** 一般而言 M-K 方法计算的趋势符合正态分布, 且  $Z = 4.381 > 1.96$ , 则说明该点海表温度的日变化趋势有明显上升趋势。UFK 和 UBK 在 95% 的信度线之间没有交点, 说明该点温度在这段时间内没有突变。

### 3.2.2 研究点海表温度的年变化趋势与突变情况

#### ①趋势变化:

表 2 海表年温度变化数据趋势表

n	No significant trend		Sens Nonparametric Estimator		
	Mean Value	Z statistic	斜率估计	置信下限	置信上限
9	29.6394	0.31277	0.010533	-0.099096	0.092132

#### ②突变情况:

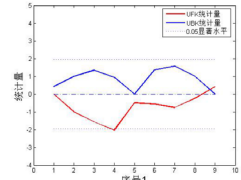


图 4 海表年温度突变情况图

**③结果分析:**  $Z \text{ statistic} = 0.31277 < 1.96$ , 说明该点温度的年变化趋势有上升但不明显。UFK 和 UBK 在 95% 的信度线之间有交点, 且致在 8-9 年之间, 说明该点海表温度在第八年左右发生突变, 且该点表温度由下降趋势转为上升趋势。

## 3.3 周期成分识别

**3.3.1 周期显著性判断。**首先我们用方差分析法判断时间序列中是否存在显著周期。在分析周期之前, 事先并不知道这一序

列的周期是多少，所以要根据序列长度，列出可能存在的周期。  
若时间序列总长度为  $n$ ，则可能存在的周期为  $T(T=2,3,\dots,n/2)$ 。将  
周期为  $T$  的时间序列进行分组，对应可以分为组数据，根据

$$Z(i) = Z(i + jT) \quad (j=1,2,3,\dots)$$

可知第  $i$  组的平均值为：

$$\bar{Z}_i = \frac{1}{m} \sum_{j=0}^{m-1} Z(i + jT) \quad (i=1,2,3,\dots,T)$$

上式中  $m$  表示第  $i$  组内的数据个数。

将时间序列分别按每一个可能的周期  $T(T=2,3,\dots,n/2)$  进行分  
组，计算相应的组内离差平方和  $S_E$  及组间离差平方和  $S_A$ ：

$$S_E = \sum_{i=1}^T \sum_{j=1}^m (Z_{ij} - \bar{Z}_i)^2 \quad (\text{自由度 } f_1 = n - T)$$

$$S_A = \sum_{i=1}^T m(\bar{Z}_i - \bar{Z})^2 \quad (\text{自由度 } f_2 = T - 1)$$

上式中  $\bar{Z}_i$  表示第  $i$  组的平均值， $\bar{Z}$  表示总平均值。

可以证明  $F = (S_A/f_2)/(S_E/f_1)$  服从自由度为  $(f_2, f_1)$  的  $F$  分  
布。给定显著性水平  $\alpha$ ，查分布表可以得到  $F_\alpha(f_2, f_1)$  的值，如果  
 $F > F_\alpha$ ，则此周期为显著周期。由于与真实周期相邻的值有可能  
通过显著性检验，我们取连续通过  $F$  检验的  $T$  值中使  $F$  值最大  
的对应  $T$  值作为主周期。用测试数据测试检验周期的存在和判  
定周期：

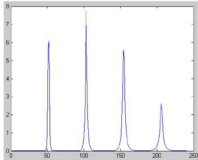


图 5 周期存在的检验和判定图

从图 5 中可以看出当  $t=52, 103, 154, 206$  时  $F$  值最大，故  
我们取定这四个周期。

3.3.2 周期提取。周期确定后，周期成份  $P_i$  可用如下形式也用  
傅里叶函数描述： $P_i = \sum_{t=1}^d (a_i \cos \frac{2\pi}{T_i} t + b_i \sin \frac{2\pi}{T_i} t)$

上式中  $d$  表示有效谐波数即主周期的个数， $T_i$  表示第  $i$  个谐  
波对应的周期， $a_i$  和  $b_i$  为参数，计算式为：

$$a_i = \frac{2}{n} \sum_{t=1}^n (Z_t \cos \frac{2\pi}{T_i} t)$$

$$b_i = \frac{2}{n} \sum_{t=1}^n (Z_t \sin \frac{2\pi}{T_i} t)$$

根据傅里叶函数求系数的公式，对其参数进行估计得到：

表 3 参数估计结果

周期	52	103	154	206
a	1.6572	0.49601	0.07085	0.021276
b	8.3538	-0.24886	0.1326	0.016165

从而得出测试数据的周期项和原始数据的对比图，从而计算  
得到随机项图，具体如下所示：

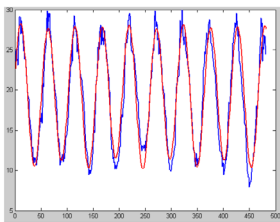


图 6 拟合周期项及原始序列

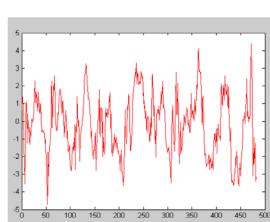


图 7 去除周期项后的随机序列

用 Eviews 对随机项做 ADF 检验，检验结果如下表所示：

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.969517	0.0000
Test critical values:		
1% level	-3.443863	
5% level	-2.867392	
10% level	-2.569950	

图 8 ADF 检验分析图

从上图中可以看出，该随机序列通过 ADF 检验，是一组平  
稳序列。再对随机序列进行相关分析，看其是否相关，从而判定  
它是否是白噪声。

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	1	0.849	0.849	348.44	0.000
2	0.740	0.067	0.132	0.000	
3	0.613	-0.109	0.795	0.000	
4	0.549	0.139	0.942	0.000	
5	0.473	-0.038	1.051	0.000	
6	0.405	-0.047	1.131	0.000	
7	0.354	0.058	1.192	0.000	
8	0.298	-0.049	1.236	0.000	
9	0.235	-0.077	1.263	0.000	
10	0.165	-0.051	1.276	0.000	
11	0.115	0.009	1.283	0.000	
12	0.047	-0.111	1.284	0.000	
13	0.014	0.047	1.284	0.000	
14	-0.034	-0.046	1.285	0.000	
15	-0.049	0.022	1.286	0.000	

图 9 序列相关分析图

从图 9 可以看出，该随机序列的自相关是拖尾的，而偏相关  
是一阶截尾的，故可建立 AR(3) 模型，模型建立结果和残差  
序列如下所示：

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.190082	0.253595	-0.749549	0.4539
AR(1)	0.797738	0.045687	17.46112	0.0000
AR(2)	0.161039	0.058202	2.766910	0.0059
AR(3)	-0.113675	0.045751	-2.484640	0.0133

图 10 随机序列模型结果分析图

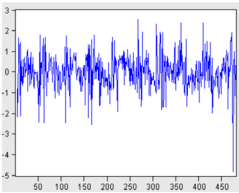


图 11 残差序列图

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.013	0.013	0.009	0.776	
2	-0.010	-0.010	0.129	0.927	
3	-0.119	-0.119	0.809	0.073	
4	0.088	0.092	1.079	0.030	
5	0.043	0.039	1.181	0.041	
6	-0.016	-0.021	1.179	0.068	
7	0.038	0.063	1.247	0.087	
8	0.081	0.083	1.295	0.075	
9	0.035	0.020	1.491	0.094	
10	-0.040	-0.027	1.689	0.109	
11	0.045	0.057	1.680	0.116	
12	-0.073	-0.087	1.938	0.081	
13	0.025	0.012	1.960	0.105	
14	-0.085	-0.072	2.140	0.058	
15	0.026	-0.001	2.347	0.075	

图 12 残差序列的序列相关图

由图 11 和图 12 可知，该残差序列在 5% 的置信区间内平稳  
且不相关，故为白噪声序列。

**总结：**本文通过 SPSS 初步检验出所处理的温度数据满足正态性。  
继而通过 Mann-Kendall 秩次相关检验发现东海表层温度的年变  
化趋势虽然在第八年有突变，海表温度由下降趋势转为上升趋势，  
但近 9 年整体无明显变化的趋势。而在日变化趋势上则有明显的  
上升趋势。周期存在性检验中， $t=52, 103, 154, 206$  时  $F$  值最大，  
故我们取定这四个周期。周期成份可用傅里叶函数来描述，4 个  
周期下系数则可分别通过参数估计得到。测试数据的周期项和原  
始数据对比图表明所得结果有效。该随机序列的自相关是拖尾的，  
而偏相关是一阶截尾的，故可建立 AR(3) 模型，分析得序列  
是白噪声序列。

作者简介：韩心娅（1994-），女，汉族，浙江杭州人。本科在读，  
单位：杭州电子科技大学理学院，研究方向：数据  
挖掘、金融数学。  
程宗毛（1964.06-），男，汉族，职称：博士、副教授，  
杭州电子科技大学应用数学专业硕士生导师，研究  
方向：保险精算、金融数学、概率统计。