

基于非平稳时间序列的不同趋势项提取方法对比研究^{* 1}

曹 奇 岳东杰 高永攀 王性猛

(河海大学测绘科学与工程系,南京 210098)

摘 要 基于时间序列分析模型,运用高阶差分、灰色模型、频谱分析和奇异谱分析方法对非平稳时间序列提取趋势项,然后对平稳残差序列进行时序建模。通过对大坝位移监测序列建模分析进行预测,发现奇异谱分析和频谱分析时序分析模型相对于传统的高阶差分和灰色模型时序分析模型的预测精度有所提高。

关键词 趋势项提取;高阶差分;灰色模型;频谱分析;奇异谱分析

中图分类号:P207

文献标识码:A

CONTRAST STUDY ON VARIOUS METHODS EXTRACTING TREND EXTRACTION BASED ON NON-STATIONARY TIME SERIES

Cao Qi, Yue Dongjie, Gao Yongpan and Wang Xingmeng

(Department of Surveying and Mapping Science and Engineering, Hohai University, Nanjing 210098)

Abstract On the basis of the time series analysis model, higher difference, gray model, spectrum analysis and singular spectrum analysis method are used to extract non-stationary time series trend, and model smooth residuals for time series. The analysis results of dam displacement monitoring series indicate that prediction accuracy is improved by using singular spectrum analysis and spectrum analysis of time series analysis model compared to the traditional model of higher difference timing analysis model and gray prediction model.

Key words: trend extraction; higher difference; gray model; spectrum analysis; singular spectrum analysis

1 引言

通常对大坝等建筑物位移量的预测模型有两种:基于高阶差分的时间序列模型预测和灰色模型预测法。其中高阶差分时间序列比较适合短周期预报,灰色模型适合长周期预报^[1]。但是,高阶差分时序分析法会造成残差信息量的浪费,导致预测精度不理想;单一的灰色模型方法只是研究了位移序列的趋势模型,没有对平稳残差序列建模。

时间序列模型的优点在于它考虑了观测数据在时间序列上的依存性,也考虑到了随机波动的干

扰^[2]。所以,时间序列模型及其分析方法在测绘领域已经有广泛的应用,考虑到大坝等建筑物的正垂径向位移量可以看作是随时间变化的动态数列,且这些数据之间存在一定的時間相关性,因此可以将时间序列分析方法用于大坝等建筑物位移量的分析与预测。

本文采用灰色模型、频谱分析和基于奇异谱分析方法对位移时间序列进行趋势项提取,利用 AR-MA 模型对去除趋势项的序列建模,即在原来的位移模型上加上随机补偿,使原来模型更加完善,提高预报的精度;对高阶差分时序法、灰色模型时序法、

* 收稿日期:2013-07-07

作者简介:曹奇,男,1990年生,硕士研究生,研究方向:大地测量与测量工程. E-mail:553140514@qq.com

频谱分析法和奇异谱分析时序法的预测模型进行精度比较。

2 基于不同趋势项提取的时间序列分析模型

若将非平稳序列的样本观测值记作 X_1, X_2, \dots, X_N , 对应的时间序列 $\{X_t\}$ 的模型为:

$$X_t = \mu_t + Y_t \quad (1)$$

式中 μ_t 表示 X_t 中随时间变化的均值, Y_t 是 X_t 中剔除趋势项或周期项 μ_t 后的剩余部分。

非平稳时间序列建模分为直接剔除法和趋势项提取法。

直接剔除法。定义差分算子 ∇ , 则 $\nabla X_t = X_t - X_{t-1}$ 称对 X_t 的一阶差分^[3], 以此类推, d 阶差分公式为:

$$\nabla^d X_t = X_t - C_d^1 X_{t-1} + C_d^2 X_{t-2} + \dots + (-1)^{d-1} C_d^{d-1} X_{t-d+1} + (-1)^d X_{t-d} \quad (2)$$

采用灰色模型、频谱分析和奇异谱分析方法来提取序列的趋势项。

1) GM(1,1) 灰色模型

已知原始非负序列 $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\}$, 一次正向累加生成新序列 $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)\}$, 其中:

$$x^{(1)}(t) = \sum_{i=1}^t x^{(0)}(i), (t=1, 2, \dots, n) \quad (3)$$

建立灰色 GM(1,1) 模型为^[4]

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \quad (4)$$

估值序列为

$$\hat{x}^{(0)}(t+1) = (1 - e^a)(x^{(0)}(1) - u/a)e^{-at} \quad (5)$$

2) 频谱分析模型

频谱分析一般是根据原始序列的大致趋势用多项式函数(阶数 $N < 5$)拟合^[4], 利用能量函数和快速傅里叶变换函数对残差序列进行分析, 作出幅度-频率图, 找出幅度最高点, 拟合出傅里叶三角函数。

3) 奇异谱分析方法

对给定的时间序列 x_1, x_2, \dots, x_N , 给定嵌套维数 $L, L < N/2$, 建立时滞排列矩阵为^[5]

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_{N-L+1} \\ x_2 & x_3 & \dots & x_{N-L+2} \\ \dots & \dots & \dots & \dots \\ x_L & x_{L+1} & \dots & x_N \end{bmatrix} \quad (6)$$

对时间经验正交函数展开(TEOFs), 得滞后协方差为

$$S = \begin{bmatrix} s_{(0)} & s_{(1)} & \dots & s_{(L-1)} \\ s_{(1)} & s_{(0)} & \dots & s_{(L-2)} \\ \dots & \dots & \dots & \dots \\ s_{(L-1)} & s_{(L-2)} & \dots & s_0 \end{bmatrix} \quad (7)$$

S 为对称矩阵且主对角线为同一常数(Toeplitz 矩阵), 求解 S 特征值和特征向量等价于对 X 矩阵作 TEOFs, 其全体特征值

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0 \quad (8)$$

即为序列 $\{x_t\}$ 的奇异谱, 对其奇异值进行运算称为奇异谱分析^[6-8]。 λ_k 对应的特征向量为第 k 阶模式, λ_k 对应特征向量 E_k 称为时间经验正交函数(TEOF), 第 k 个时间主分量(TPC)定义为原序列 $\{x_t\}$ 在 E_k 上正交投影系数, 即

$$a_{ik} = \sum_{j=1}^M x_{i+j} E_{kj} \quad 0 \leq i \leq N-M, 1 \leq j \leq M \quad (9)$$

式中 a_{ik} 是 E_{kj} 表示的时间型在原序列 $x_{i+1}, x_{i+2}, \dots, x_{i+m}$ 时段的权重。

SSA 的重要功能由重建成分(RC, Reconstruction component)实现^[9,10]。所谓的重建是由 TEOF 和 TPC 重建一个长度为 N 的序列, 第 k 个重建 x_i 的成分记 x_{ki} , 公式为

$$x_{ki} = \begin{cases} \frac{1}{i} \sum_j a_{i-j,k} E_{kj} & 1 \leq i \leq M-1 \\ \frac{1}{M} \sum_{j=1}^M a_{i-j,k} E_{kj} & M \leq i \leq M+1 \\ \frac{1}{N-i+1} \sum_{j=i-N+L}^i a_{i-j,k} E_{kj} & N-M+2 \leq i \leq N \end{cases} \quad (10)$$

RC 具有叠加性质, 对所有 RC 求和得原序列为

$$x_i = \sum_{k=1}^M x_{ki} \quad (11)$$

3 趋势项提取模型处理流程

趋势项提取模型处理流程为:

1) 对获取的位移时间序列根据莱茵达法则进行粗差探测, 剔除异常值得到序列 $X_t^{(0)}$;

2) 对序列 $X_t^{(0)}$ 利用高阶差分、灰色模型、频谱分析和奇异谱分析进行趋势项剔除或提取, 得到残差序列 $X_t^{(1)}$, 计算残差均方根误差, 检验模型拟合精度;

3) 对残差序列 $X_t^{(1)}$ 均值化处理生成 $X_t^{(2)}$, 计算其自相关系数和偏相关系数, 初步建立 ARMA(p, q) 模型;

4) 用最小信息准则(AIC)判定时间序列模型的阶数^[1], 对其时间序列模型用 F 检验准则和残差平方和最小准则来检验模型适应性;

5) 若模型通过适应性检验则对序列拟合预报,

再根据式(1)对序列进行还原,对位移时间序列模型预测。

4 大坝监测位移时间序列模型分析

以某大坝一坝段的正垂径向位移的观测数据序列(表1)为例,取其前36期数据进行模型分析,用前30期数据为基础建模,对后面6期数据进行预测分析。

表1 坝段正垂径向位移观测时间序列(单位:mm)
Tab.1 Observation time series of vertical radial displacement of the dam (unit:mm)

期数	位移	期数	位移	期数	位移	期数	位移	期数	位移
1	5.95	7	4.78	13	4.29	19	3.55	25	2.78
2	5.96	8	4.47	14	4.22	20	2.00	26	2.67
3	5.74	9	3.69	15	4.31	21	1.77	27	2.31
4	5.13	10	3.88	16	4.39	22	1.78	28	2.61
5	5.03	11	4.87	17	4.28	23	1.99	29	2.44
6	4.89	12	4.49	18	3.79	24	2.33	30	1.70
								31	1.76
								32	1.35
								33	1.47
								34	1.89
								35	2.02
								36	2.37

4.1 高阶差分时序分析模型

30期位移数据进行预处理,具有明显的下降趋势,对其进行二阶差分处理,去除趋势项。经自相关系数与偏相关系数和AIC最小值准则判断确定为ARMA(6,5)模型:

$$X_t = \varphi(1,1)X_{t-1} + \varphi(2,1)X_{t-2} + \varphi(3,1)X_{t-3} + \varphi(4,1)X_{t-4} + \varphi(5,1)X_{t-5} + \varphi(6,1)X_{t-6} + a_t - \theta(1,1)a_{t-1} - \theta(2,1)a_{t-2} - \theta(3,1)a_{t-3} - \theta(4,1)a_{t-4} - \theta(5,1)a_{t-5}$$

其中{a_t}白噪声序列,满足:

$$E(a_t) = 0, Var(a_t) = \sigma^2,$$
$$\varphi = \begin{bmatrix} -0.677\ 3 & -0.076\ 2 & -0.089\ 0 & -0.105\ 7 & 0.255\ 6 & 0.015\ 4 \end{bmatrix}^T,$$
$$\theta = \begin{bmatrix} -0.453\ 0 & 0.110\ 9 & -0.121\ 2 & 0.533\ 6 & -0.902\ 7 \end{bmatrix}^T.$$

4.2 灰色模型时序分析模型

对30期位移序列用灰色模型进行趋势项提取(图1), $e^{-0.036\ 6t}$ 最小二乘算法求解出微分方程,得到

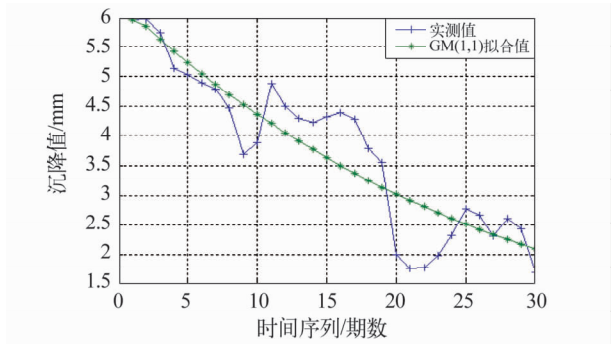


图1 GM(1,1)模型提取趋势项
Fig.1 Trend items extracted by GM(1,1) model

估值序列: $\hat{x}^{(0)}(t+1) = -162.421\ 6(1 - e^{0.036\ 6})e^{-0.036\ 6t}$ 。其小误差概率 $P=0.827\ 6$,后验方差 $C=0.180\ 5$,说明模型拟合精度比较高。

对残差时间序列ARMA模型建模,确定为AR(2)模型,用最小二乘法计算得到预测公式为 $X_t = 1.131\ 8X_{t-1} - 0.456\ 5X_{t-2} + a_t$,其中{a_t}白噪声序列,满足 $E(a_t) = 0, Var(a_t) = \sigma^2$ 。

4.3 频谱分析时序模型

因为原位移序列有明显的趋向性,离散程度不是很高,采用三阶函数进行大致拟合。对多项式拟合的残差序列进行傅里叶变换,得到的幅频图如图2和图3。

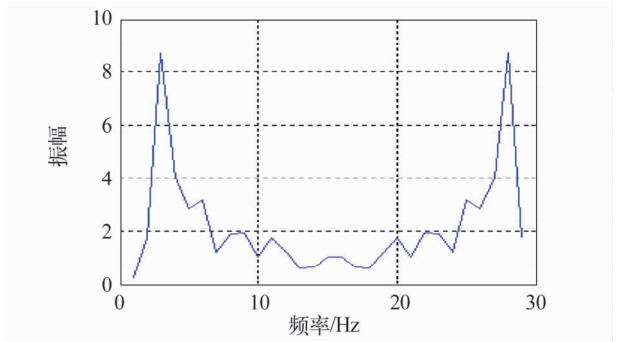


图2 残差序列傅里叶变换后的幅频图(N=30)
Fig.2 Amplitude-frequency figure of residual series after Fourier transform

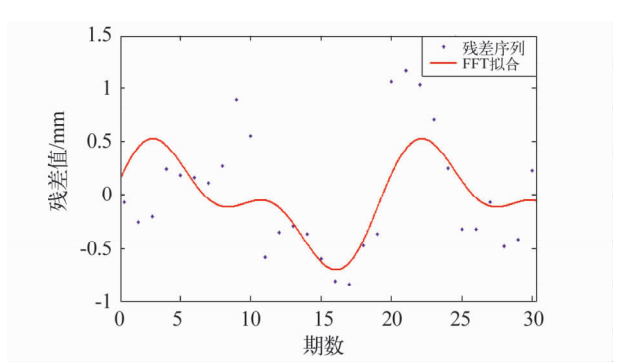


图3 傅里叶三角函数拟合曲线图
Fig.3 Curve fitted by Fourier trigonometric

用FFT对信号做谱分析,整个频谱图是以Nyquist频率为对称轴,且可以明显识别出信号中含有两种频率成分(3 Hz和6 Hz)。结合频谱分析法三角函数参数的求取,得到拟合函数为 $y = a + b\sin(2 \times \pi \times 3 \times t) + c^* \sin(2 \times \pi \times 6 \times t)$,其中 $a = -0.087\ 6, b = 0.429\ 6, c = 0.277\ 0$ 。

对模型拟合度检验:和方差为 $3.407\ \text{mm}^2$,均方根误差为 $0.393\ 7\ \text{mm}$,说明拟合精度较好。对残余时间序列时序分析,确定为AR(1)模型,用最小二乘法计算得到预测公式为 $X_t = 0.521\ 3X_{t-1} + a_t$,其中{a_t}为白噪声序列,满足 $E(a_t) = 0, Var(a_t) = \sigma^2$ 。

4.4 奇异谱分析时序模型

对窗口长度 $L=6$ 进行计算,根据曲率谱最大峰值准则,有效奇异值个数取 1,选第一层为有效主分量。用第一时间主成分进行重建,利用重建序列与原位移时间序列的均方误差选择窗口长度。

其均方根误差为 0.426 1 mm,说明 SSA 提取的趋势项效果较好,有较高的模型拟合精度。对 RC1 重建序列用线性递推公式(LRF)从 31 到 36 进行递推延伸^[5,6],图 4 为 LRF 算法对 RC1 的预测序列。

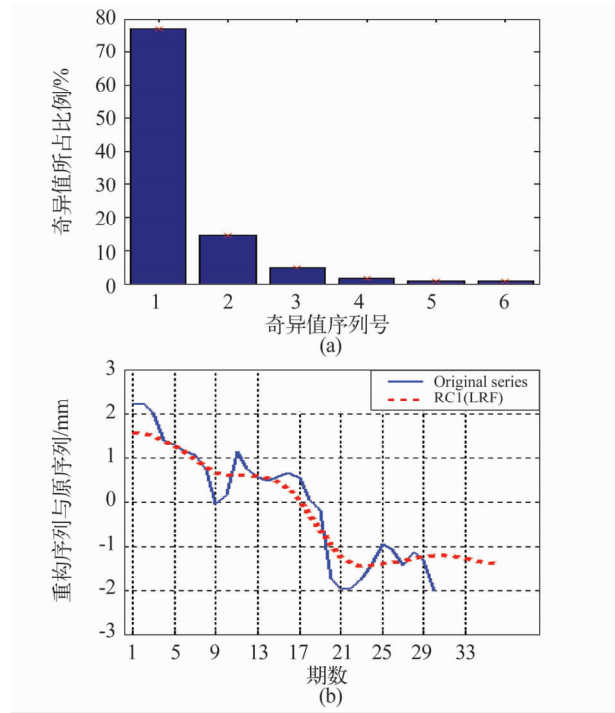


图 4 窗口长度为 6 的奇异值所占比例和其第一主分量重构序列图

Fig.4 Eigenvalue variance and the first main component of the reconstructed time series($L=6$)

对残差时间序列时序分析,确定为 AR(2) 模型,用最小二乘法计算得到预测公式为 $X_t=0.753\ 9\ X_{t-1}-0.322\ 4X_{t-2}+a_t$, $\{a_t\}$ 为白噪声序列,满足 $E(a_t)=0,Var(a_t)=\sigma^2$ 。

4.5 模型预测精度对比分析

表 2 为不同模型对位移序列的 6 期预测量;表 3 为不同预测模型的残差准确率和均方根误差对比,残差准确率定义为 $(1-残差/预测值)\%$ 。从表 2、3 可以得出:奇异谱分析时序法和频谱分析时序法相对于传统的高阶差分时序法和灰色模型时序法的预测精度有提高;奇异谱时序法预测序列波动性很小,对于突变值的预测效果较差;频谱分析时序法预测精度较高,但对序列的平稳性要求较高。模型精度提高有点缓慢,可能存在以下原因:原始数据序列步长较短且基数较小;数据序列预测的一定范围

表 2 不同模型对位移时间序列预测量(单位:mm)

Tab.2 Predicted displacement time series with various models (unit:mm)

预测模型	高阶差分		灰色模型		频谱分析		奇异谱分析	
	时序法	残差	时序法	残差	时序法	残差	时序法	残差
期数	实测值	预测值	残差	预测值	残差	预测值	残差	预测值
31	1.76	2.58	-0.82	1.36	0.40	1.56	0.20	1.91
32	1.35	2.59	-1.24	1.13	0.22	1.27	0.08	2.29
33	1.47	1.66	-0.19	1.19	0.28	1.78	-0.31	2.42
34	1.89	2.73	-0.84	1.34	0.55	2.13	-0.24	2.32
35	2.02	3.06	-1.04	1.42	0.60	2.46	-0.44	2.19
36	2.37	3.01	-0.64	1.50	0.87	2.50	-0.13	2.22

表 3 不同预测模型序列与原序列的残差准确率和均方根误差对比(单位:%)

Tab.3 Residual accuracy with various prediction models and RMSE contrast (unit:%)

期数	高阶差分	灰色模型	频谱分析	奇异谱分析
	时序法	时序法	时序法	时序法
31	68.22	70.15	87.47	92.08
32	52.17	80.14	93.43	58.97
33	88.79	76.76	82.68	60.82
34	69.16	58.45	88.54	81.64
35	66.10	57.79	81.96	92.35
36	82.27	58.00	94.80	93.24
均方根	0.898 1	0.435 7	0.280 6	0.436 0
(mm)				

内会自然振动,具有偶然性;数据预处理和模型预测还需要优化设计。

5 结论

1)灰色模型时序法、频谱分析时序法与奇异谱分析时序法相对于传统的高阶差分时序分析法预测精度有显著提高,说明对趋势项的有效提取结合对平稳残差序列建立 ARMA 模型能有效提高拟合与预测的精度;

2)频谱分析时序模型从时域转换到频域内进行特征分析,时频域相结合,具有很强预测诊断能力;

3)奇异谱分析提取趋势项后采用 LRF 算法延展趋势项结合时序模型能有效提高预测精度,尤其对于稳定平稳的位移时间序列。

参 考 文 献

1 杨位钦,顾岚. 时间序列分析与动态数据建模[M]. 北京:北京工业学院出版社,1986. (Yang Weiqin and Gu lan. Time series analysis and dynamic data modeling[M]. Beijing:Beijing Institute of Technology Press,1986)

2 李晓宇,等. 基于改进灰色 ARMA 模型的卫星钟差短期预报研究[J]. 大地测量与地球动力学,2013,(1):59-63. (Li Xiaoyu, et al. Research on short-term prediction of

satellite clock errors based on improved grey ARMA model [J]. Journal of Geodesy and Geodynamics, 2013, (1): 59 – 63)

3 隋铭明, 陈健, 史玉峰. 时间序列分析与频谱分析联合用于变形监测分析与预报[J]. 工程勘察, 2011, 39(11): 77 – 80. (Sui Minming, Chen Jian and Shi Yufeng. Integration of time series analysis and spectrum analysis and its application in deformation data processing and forecasting [J]. Geotechnical Investigation and Surveying, 2011, 39(11): 77 – 80)

4 陈启华, 等. 基于抗差估计的灰色模型与时间序列组合模型及其在变形预测中的应用[J]. 工程勘察, 2012, 40(8): 51 – 54. (Chen Qihua, et al. Combination of the gray model with time-series model based on robust estimation and its application in the deformation prediction [J]. Geotechnical Investigation and Surveying. 2012, 40(8): 51 – 54)

5 Vautard Robert and Pascal Yiou Michael Ghil. Singular-spectrum analysisa toolkit for short noisy chaotic signals [J]. Physical(D), 1992, 58: 95.

6 Broomhead D S and King P. Extracting qualitative dynamics for experimental data[J]. Physica(D), 1987, 20: 217.

7 江志红, 丁裕国. 奇异谱分析的广义性及其应用特色[J]. 气象学报, 1998, (6): 97 – 106. (Jiang Zhihong and Ding Yuguo. Generality and applied features for singular spectrum analysis [J]. Journal of Applied Meteorological Science, 1998, (6): 97 – 106)

8 Shen Yunzhong and Li Weiwei. Spatio temporal related signal and noise analysis of GPS monitoring series of the base stations in China[R]. International Union of Geodesy and Geophysics Conference, 2012.

9 王解先, 连丽珍, 沈云中. 奇异谱分析在 GPS 站坐标监测序列分析中的应用[J]. 同济大学学报(自然科学版), 2013, (2): 282 – 288. (Wang Jiexian, Lian Lizhen and Shen Yunzhong. Application of singular spectral analysis to GPS station coordinate monitoring series [J]. Journal of Tongji University for Natural Science, 2013, (2): 282 – 288)

10 吴洪宝. 奇异谱分析—最大熵预报方法[J]. 甘肃气象, 2000, (1): 1 – 5. (Wu Hongbao. Singular spectral analysis—maximum entropy forecast method [J]. Gansu Meteorology, 2000, (1): 1 – 5)

(上接第 149 页)

Journal of Arid Environments, 1998, 39(2): 143 – 154.

4 董得红. 大比例尺 TM 卫星影像图在全省沙漠化土地普查与监测中的应用[J]. 青海环境, 1996, 6(3): 118 – 120. (Dong Dehong. Application of large scale TM satellite image to the exploration and monitoring of the desertification land [J]. Journal of Qinghai Environment, 1996, 6(3): 118 – 120)

5 王涛, 等. 中国北方沙漠化土地时空演变分析[J]. 中国沙漠, 2003, 23(3): 230 – 235. (Wang Tao, et al. Time-space evolution of desertification land in northern China [J]. Journal of Desert Research, 2003, 23(3): 230 – 235)

6 袁佩新, 周明伟. 西藏自治区土地沙漠化遥感调查[J]. 四川地质学报, 2004, 24(4): 223 – 225. (Yuan Peixin and Zhou Mingwei. Invastigation of desertification land in Tibet with remote sensing [J]. Journal of Sichuan Geology, 2004, 24(4): 223 – 225)

7 卫亚星, 王莉雯, 王一谋. 利用 TM 监测中国西部沙漠化进程[J]. 水土保持通报, 2004, 24(4): 47 – 50. (Wei Yaxing, Wang Liwen and Wang Yimou. Monitoring the desertification in west China using TM [J]. Bulletin of Soil and Water Conservation, 2004, 24(4): 47 – 50)

8 李爱敏, 等. 21 世纪初科尔沁沙地沙漠化土地变化趋势[J]. 地理学报, 2006, 61(9): 976 – 984. (Li Aimin, et al. Development trend of desertification in Korqin in the earlier 21st century [J]. Acta Geographica Sinica, 2006, 61(9): 976 – 984)

9 杨婷婷, 等. 基于 TM 影像的镶黄旗北部沙地沙漠化动态监测[J]. 西北林学院学报, 2009, 2: 145 – 147. (Yang Tingting, et al. Monitoring desertification in the north Xiang-huang Banner using TM [J]. Journal of Northwest Forestry University, 2009, 2: 145 – 147)

10 李辉霞, 等. 基于 NDWI 的海南岛西部沙漠化信息自动提取方法研究[J]. 中国沙漠, 2006, 26(2): 215 – 219. (Li Huixia, et al. A method extractiing desertification information in western Hainan island based on NDWI [J]. Journal of Desert Research, 2006, 26(2): 215 – 219)

11 曾永年, 等. 基于地表定量参数的沙漠化遥感监测方法[J]. 国土资源遥感, 2005, 2: 40 – 44. (Ceng Yongnian, et al. A method to detect desertification based on remote sensing synthetic index [J]. Remote Sensing for Land & Resources, 2005, 2: 40 – 44)

12 王晓怡, 张德强, 姚磊. TM 遥感影像信息及最佳波段组合研究[J]. 山东师范大学学报(自然科学版), 2010, 25(1): 109 – 112. (Wang Xiaoyi, Zhang Deqiang and Yao Lei. Study on thematic mapper data and the optimal band combination [J]. Journal of Shandong of Normal University (Natural Science), 2010, 25(1): 109 – 112)