

海平面变化的小波和自回归模型集成预测试验

袁林旺, 谢志仁, 钟鹤翔

(南京师范大学 地理科学学院, 江苏 南京 210097)

摘要: 针对海面变化预测时间序列模型中趋势组份和周期(准周期)组份的提取和预测问题, 基于吴淞站 1955~2001 年月平均潮位序列, 采用小波分析(WA)与自回归(AR)模型相结合的方案, 对小波分解的不同尺度分量序列, 借助于时间序列模型进行分量预测, 再对它们进行叠加建立预测模型, 进行了月平均潮位预测试验。以 1955~1996 年数据为基础建立模型, 1997~2001 年数据作为验证, 结果表明两种方法的结合使用显示了较好的效果, 具有较高的精度。

关键词: 海平面变化; 预测; 小波分析(WA); 自回归(AR)模型

中图分类号: P731.34

文献标识码: A

文章编号: 1000-3096(2008)04-0031-06

海平面变化趋势预测是一个跨学科的、难度很大的研究课题。依据对验潮数据的数学分析进行外推预测是一个重要的研究方向。中国沿海的验潮站数量不少, 但同时具备长序列潮位记录和地面升降精密水准测量数据的站点极少。上海吴淞站自 1912 年建立, 同时上海地区陆地水准测量与永久性水准基点的设置亦具有相当长的历史^[1]。因此吴淞站的潮位序列是中国进行海平面变化相关数学分析的理想站点之一。

作者以吴淞站经过地面沉降校正的 1955~2001 年月平均潮位序列为研究对象, 借助于小波分析方法(Wavelet Analysis, 简称 WA), 结合自回归(AR)模型对海平面变化预测问题进行了探讨。作者侧重于方法学讨论, 以 1955~1996 年数据为基础建立模型, 以 1997~2001 年的数据验证预测的可信度。

1 潮位序列预测海面变化的基本思路

在早期研究中, 多采用线性方程来表示海平面变化的变率。由于线性方程预测与数据序列的长度与序列的时段密切相关, 其预测结果仅能提供粗线条的“定性”描述。因此除了用于确定大致趋势外, 后续的研究往往均考虑了周期项等其他成分, 并对趋势项的函数形式进行了探讨, 在方法上也进行了多种尝试。

黄立人^[2]采用线性方程拟合趋势项, 对去除趋势项后的序列通过谱分析确定显著周期, 建立上述

不同显著周期所对应的方程, 构建海平面变化的累加预报模型。左军成等^[3]建立了思路上大致类似的本征函数预报模型。郑大伟等^[4]采用跳步时间序列分析模型对吴淞站月平均潮位序列进行了预测, 并借助于小波分析方法获得了序列变化的几个主导周期。秦曾灏等^[5]采用非线性回归模型来确定趋势项。左军成^[6]对月平均海面序列进行一次累加, 建立了 GM(1, 1)灰色预测模型。夏华永^[7]认为灰色模型对存在显著周期变化的月平均海面预测时效果一般。因此提出了带周期项的灰色分析模型。

综上所述, 平均海平面变化的主体部分可表示为趋势项和周期项的叠加, 即将月(年)平均海面序列 $Y^{(0)}(t)$ 表示成如下形式:

$$Y^{(0)}(t) = T^{(0)}(t) + P^{(0)}(t) + X^{(0)}(t) + a^{(0)}(t) \quad (1)$$

式中, $T^{(0)}(t)$ 为确定性的趋势项; $P^{(0)}(t)$ 为确定性周期项; $X^{(0)}(t)$ 为剩余随机序列; $a^{(0)}(t)$ 为白噪声序列。

尽管诸多研究已对海平面变化预测模型进行了

收稿日期: 2004-11-24; 修回日期: 2007-11-09

基金项目: 国家自然科学基金资助项目(40171008); 国家 973 计划前期研究专项(2007CB416602)

作者简介: 袁林旺(1973), 男, 江苏海安人, 副教授, 博士, 主要研究方向为区域环境建模及地理信息系统, E-mail: yuanlinwang@njnu.edu.cn; 谢志仁, 通讯作者, 教授, 博士, 博士生导师, 主要研究方向为海面变化及其预测研究, E-mail: xiezhiren@njnu.edu.cn

多方面的探讨,但下述问题仍有深入探究的必要:

(1) 趋势项的形式及其对预测模型的影响;(2) 周期正弦波假设的可信性。问题(1)涉及到方程拟合趋势项对海平面变化真实趋势的逼近程度;问题(2)则与海平面变化周期的规则程度有关。

作者采用 WA 与 AR 模型相结合的方案,为上述问题的深入探讨提供一个新的案例。其基本思路是:对于月平均潮位序列采用小波分析进行分解,获得几个重要的主导周期的变化序列,借助于 AR 进行分量预测,再对它们进行叠加,最终获得序列预测值。

2 方法概述

2.1 WA

WA 是一种窗口大小(即窗口面积)固定但其形状可改变,即时间窗和频率窗都可改变的时频局部化分析方法。传统的傅立叶变换只能获得信号整体的、单层次的和单分辨率的认识。小波变换的基本思想与傅立叶变换类似,但它既可以分析信号的概貌,又可分析信号的细节,所以被誉为数学显微镜^[8,9]。小波变换最主要的优势在于它能够揭示信号的层次特征,包括尺度行为,即可在预先定义的尺度上对原始信号进行多尺度滤波^[10]。

小波变换采用信号在由一簇基函数张成的空间的投影来表征该信号。这一簇函数所具有的一个显著的特点是函数系通过一个基本母函数在不同尺度上的伸缩和平移来实现。其含义是:把某一被称为基本小波(母小波, Mother wavelet)的函数 $\varphi(t)$ 作位移 τ 后,再在不同尺度 a 下与待分析信号 $x(t)$ 作内积:

$$WT_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \phi^* \left(\frac{t-\tau}{a} \right) dt, (a > 0) \quad (2)$$

等效的频域表示是:

$$WT_x(a, \tau) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} X(\omega) \phi^*(a\omega) e^{+j\omega\tau} d\omega \quad (3)$$

式中, $X(\omega)$, $\phi(\omega)$ 分别是 $x(t)$, $\varphi(t)$ 的傅立叶变换。

2.2 AR 模型

作为随机时间序列线性模型 3 种基本形式之一

的 AR 模型在预测时,只考虑序列本身历史数据反映和包含的信息,实际上是对预测指标历史数量变动规律进行了整体概括,预测结果可信度高。该方法适用于指标数量不大,但预测频度较高的短期预测。

若序列值 x_t 是现在和过去的误差或白噪声以及先前序列值的线性组合^[11],即

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + a_t \quad (4)$$

该模型称为序列的自回归模型,简记为 AR(p) 模型, p 为自回归阶数。参数 $\varphi_i (i=1, 2, \dots, p)$ 称为自回归参数。

由于时间序列往往还存在着季节性(周期性)成分,因此在实际工作中,常采用季节性自回归模型(SAR)描述,其方程为:

$$\varphi(B)\phi(B^s) \nabla_s^D \nabla^d x_t = a_t \quad (5)$$

式中,

D 为季节差分阶数; d 为直接差分阶数;

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p;$$

$$\phi(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_m B^{ps};$$

B 为后移算子,使得 $Bx_t = x_{t-1}$, $B^m x_t = x_{t-m}$; s 为季节(周期)长度。

3 预测试验的基本步骤

通过小波变换获得不同尺度(周期)上的潮位变化信息,对不同尺度潮位变化分别采用时间序列模型进行预测,并对各分量预测结果进行加和,则可以获得在不同尺度因素作用下的潮位变化趋势。

3.1 基于 WA 的分量提取

1988 年 Daubechies 基于 Mallet 的多分辨率思想,构造了 Daubechies 小波系(在 MATLAB 中简记为 db N ,其中 N 为小波序号)。该小波系具有正交性及紧支集,且仅在一有限区域中是非零的二进制小波,这些特点使其成为应用最为广泛小波母函数之一^[8]。

取吴淞验潮站 1955 年 1 月~ 2001 年 12 月的月平均潮位序列采用 db5 小波进行了 4 层次的离散小波分解,并采用对称延伸法消除了边界效应影响,其分解结果如图 1。对所获得的低频成分及 4 个尺度进行谱分析可知,低频成分代表 7 a 左右的周期(图 1B);

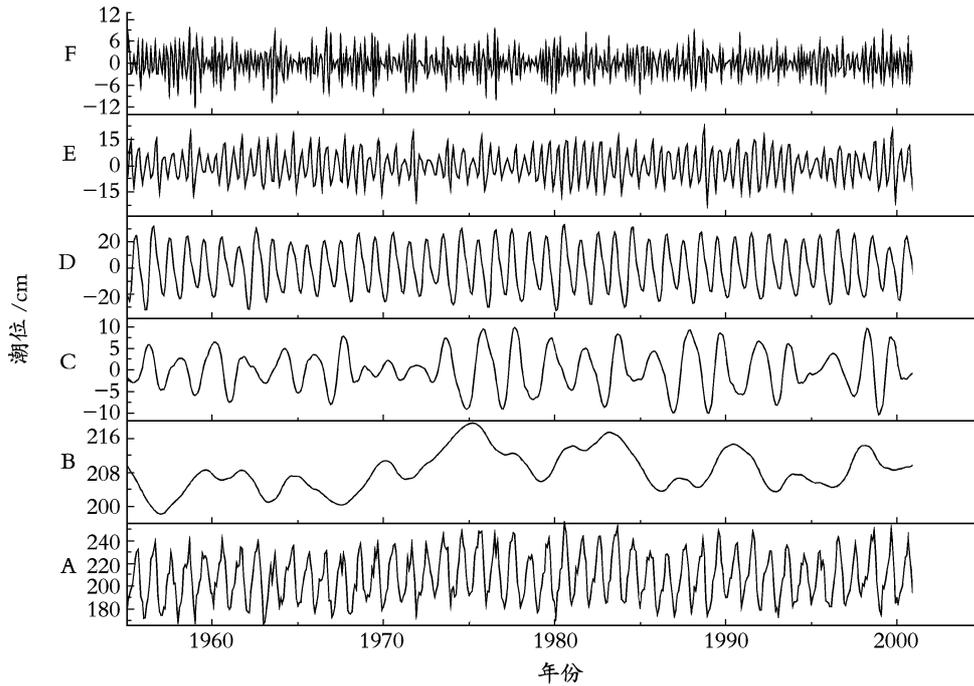


图 1 小波分解结果

Fig. 1 Results of wavelet decomposition

A. 原始序列; B. 低频成分; C, D, E, F. 分别为尺度 4, 3, 2, 1 分解结果

A. original serial; B. lower frequency element; C, D, E, F. are the decomposition results of scale 4, 3, 2, 1 respectively

尺度 4 分解结果代表 2 a 左右的周期成分(图 1C); 尺度 3 代表年周期(图 1D); 尺度 2 分解结果代表 6 个月的周期(图 1E); 尺度 1 分解结果代表更高分率的成分(图 1F)。

3.2 基于 SAR 的预测

在小波分解的多个尺度的序列中, 高频部分(尺度 1)的分解结果可理解为随机误差。而低频长期趋势成分表现为非平稳序列, 因此进行二阶差分处理, 以消除序列中的长期趋势, 使序列平稳化。对经平稳处理后的序列, 以及其他表现准 2 a, 1 a 周期及 6 个月周期序列, 选择 1955 年 1 月~ 1996 年 12 月吴淞站月平均潮位序列数据, 采用 SAR 分别建立模型, 模型方程分别为:

$$(1 - \varphi_1 B)(1 - \varphi_1 B^{16})x_t = a_t \quad (\text{长期趋势})$$

$$(1 - \varphi_1 B)(1 - \varphi_1 B^{12} - \varphi_2 B^{24})x_t = a_t \quad (2 \text{ a 周期})$$

$$(1 - \varphi_1 B)(1 - \varphi_1 B^{24} - \varphi_2 B^{48} - \varphi_3 B^{72} - \varphi_4 B^{96})x_t = a_t \quad (6 \text{ 个月周期})$$

$$(1 - \varphi_1 B)(1 - \varphi_1 B^{24} - \varphi_2 B^{48})x_t = a_t \quad (1 \text{ a 周期})$$

式中, B 为后移算子; a_t 为常数。

各模型参数见表 1。

表 1 不同序列的模型参数

Tab. 1 Model parameters of each serial

| 周期序列 | 模型参数 | 参数值 | 渐近标准差 | 95% 信度下界 | 95% 信度上界 |
|-------|-------------|--------|-------|----------|----------|
| 6 个月 | a_t | -0.522 | 0.618 | -2.575 | 1.532 |
| | φ_1 | 0.358 | 0.000 | 0.275 | 0.441 |
| | ϕ_1 | 0.512 | 0.000 | 0.420 | 0.604 |
| | ϕ_2 | 0.088 | 0.091 | -0.014 | 0.191 |
| | ϕ_3 | 0.231 | 0.000 | 0.126 | 0.336 |
| | ϕ_4 | 0.103 | 0.039 | 0.005 | 0.201 |
| 1 a | a_t | -6.784 | 0.004 | -11.415 | -2.153 |
| | φ_1 | 0.854 | 0.000 | 0.806 | 0.902 |
| | ϕ_1 | 0.643 | 0.000 | 0.559 | 0.726 |
| | ϕ_2 | 0.344 | 0.000 | 0.259 | 0.430 |
| 准 2 a | a_t | -0.804 | 0.269 | -2.231 | 0.624 |
| | φ_1 | 0.949 | 0.000 | 0.920 | 0.978 |
| | ϕ_1 | -0.259 | 0.000 | -0.336 | -0.181 |
| 长期趋势 | ϕ_2 | 0.488 | 0.000 | 0.411 | 0.566 |
| | a_t | 0.002 | 0.668 | -0.006 | 0.009 |
| | φ_1 | 0.793 | 0.000 | 0.739 | 0.846 |
| | ϕ_1 | -0.695 | 0.000 | -0.760 | -0.631 |

根据上述 4 个模型,即可对 1997 年 1 月~ 2001 年 12 月吴淞站月平均潮位序列进行预测,从而可通过实测数据来检验模型预测的精度。

3.3 预测结果对比

图 2 即为上述各分量 1997~ 2001 年预测结果与吴淞站实测的原始序列的对比。

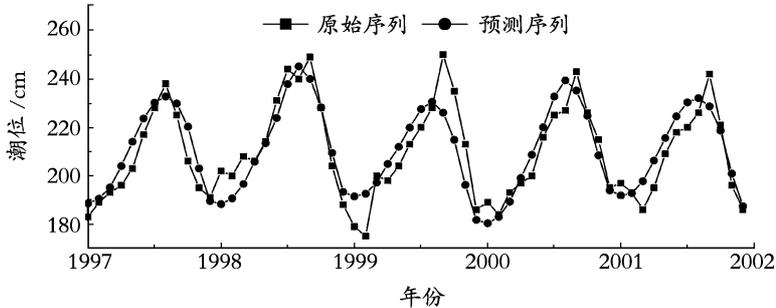


图 2 预测结果与实测数据的对比

Fig. 2 Correlation between prediction result and original serial

结果表明,两条曲线之间总体上表现出了相当好的吻合度。吻合度较差的点多数落在 1998 年和 1999 年,这与这两年月平均潮位变化表现出不同的规律性有关。从图 2 可以看出,1998 年头 3 个月的月平均潮位较其他各年偏高,且当年 9 月份出现了 1955~ 2001 年全序列中第三高(最大值为 252 cm, 1983 年 9 月)的月平均潮位值(249 cm);而 1999 年头 2 个月出现了异常低潮位,接着在 9 月份出现了全序列中的第二高度(250 cm)。这一连串异乎寻常的潮位变动的背景是 1998 年的长江大洪水、创记录的南极臭氧洞、世纪最强厄尔尼诺及其向拉尼娜的急转等一系列全球气候异常变动。作者提出的预测模型仍侧重于对序列整体规律的描述和把握,对于表现出明显异常性的变动,需要通过其他途径加以

分析。

4 预测结果的对比分析

为讨论预测的可信度,对预测序列与原始序列之间进行了绝对误差和相对误差分析。

4.1 月均序列误差分析

图 3 显示了预测的月均序列相对于原始序列的绝对误差和相对误差。由图可以看出,全部 60 个预测点(月)中有 47 个点(月)的绝对误差落在 ± 10 cm 以内,最大值出现在 1999 年 9 月,预测绝对误差的全序列平均值为 6.99 cm。而相对误差多落在 $\pm 5\%$ 以内,误差较大的几个月份同样落在 1999 年,全部预测点(月)相对误差的平均值为 3.33%。

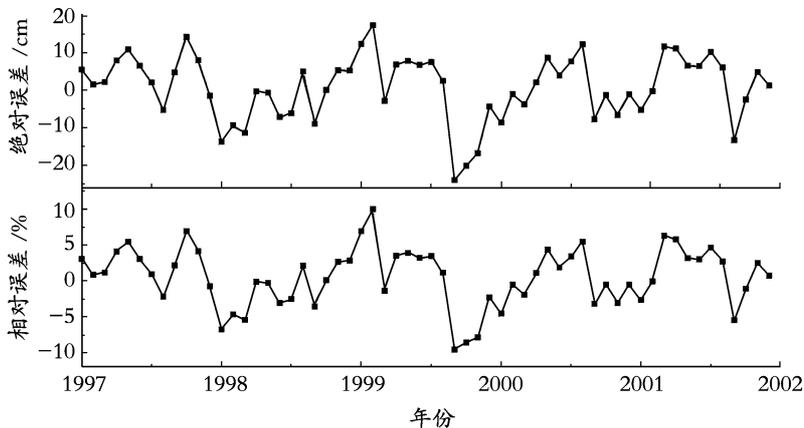


图 3 预测结果与原始序列间的绝对误差和相对误差

Fig. 3 Absolute error and relative error between prediction result and original series

4.2 年均序列误差分析

作者还针对年平均潮位的拟合结果和预测结果进行了误差分析,结果如表 2 所示。表中合成序列是前文所述小波分解的除尺度 1(代表随机变化部分)各成分的合成结果,显示 WA 提取的可预测的组份对实测数据的拟合结果。从表 2 可知,对于年均序列的预测结果与原始序列之间的绝对误差均小于 5 cm,相对误差均小于 3%,而 1999 年和 2000 年预测的绝对误差小于 1 cm;而相对误差小于 1%。见表 2 年均序列值、绝对误差和相对误差对比

Tab.2 Correlation between annual average value, absolute error and relative error

| 年份 | 序列值(cm) | | | 绝对误差(cm) | | | 相对误差(%) | | |
|------|---------|----------|---------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 原始序列 | SSA 合成序列 | 预测序列 | 合成序列与原始序列 | 预测序列与原始序列 | 预测序列与合成序列 | 合成序列与原始序列 | 预测序列与原始序列 | 预测序列与合成序列 |
| 1997 | 205.333 | 205.494 | 210.12 | 0.16 | 4.787 | 4.626 | 0.1 | 2.3 | 2.3 |
| 1998 | 217.833 | 217.681 | 214.352 | -0.152 | -3.481 | -3.329 | -0.1 | -1.6 | -1.5 |
| 1999 | 208.417 | 208.548 | 207.873 | 0.131 | -0.544 | -0.675 | 0.1 | -0.3 | -0.3 |
| 2000 | 209.167 | 209.143 | 209.561 | -0.024 | 0.394 | 0.418 | 0 | 0.2 | 0.2 |
| 2001 | 207.417 | 207.08 | 210.525 | -0.337 | 3.108 | 3.445 | -0.2 | 1.5 | 1.7 |

5 结果与讨论

本研究表明,采用 WA 与 AR 模型预测模型相结合的方案,即应用 WA 进行多尺度分解,对各分量分别建立 AR 模型进行预测具有相当高的精度。在进行预测试验的 5 a(1997~2001 年)中,全部 60 个预测点(月)中有 47 个点(月)的绝对误差落在 10 cm 以内,平均值为 6.99 cm。而计算的相对误差多落在 $\pm 5\%$ 以内,其平均值为 3.33%。对年均值对比可知,其绝对误差均小于 5 cm,相对误差均小于 3%,其中 1999 年和 2000 年预测的绝对误差小于 1 cm,而相对误差小于 1%。研究还表明采用月均序列进行建模预测,在此基础上获得年均序列的变化规律效果更为明显。这与获得年均值时月平均序列误差正负相消有关,也与月均序列反映了小尺度的规则周期有关。

WA 作为一种多尺度的信号分析方法,克服了传统的分析方法(如傅立叶分析)仅能获得对序列整体的、单层次的和单分辨率认识的缺陷,它能够获取对序列变化局部的、多层次的和多分辨率的认识,是资料分析、诊断和预测的有力工具。因此,采用小波分析对原始序列进行分解,可获得不同层次上的分量序列,而且其波型较为规则,增强了可预报性。在此基础上采用 AR 模型建模进行预测,阶数低,可预报性强,效果稳定。WA 所分解出的低频信号,可以视为序列变化的长期趋势,它对于真实趋势的逼近程度是其他常用方法所不具备的。该方法所提取的长期趋势分量真实可信,逼近程度高。因此,采用

样的年均序列预测精度是相当理想的。这些数据表明,采用月平均潮位变化序列来预测年平均潮位变化,其效果可能较直接采用年均序列进行预测效果为佳。这可能与月平均潮位序列各月误差在正负间具有相消性有关,此外月均序列显示出更为明显的规则周期,且对较小尺度的周期变化亦能反映,这些优点是直接采用年均序列进行预测的方法所不具备的。

WA 与 AR 集成进行海面变化预测是可行和有效的。

参考文献:

- [1] 陈西庆. 近 70 年长江口海面变化研究及其意义[J]. 地理学报, 1990, 45(4): 387-398.
- [2] 黄立人. 海面变化趋势的动态预测[J]. 海洋通报, 1992, 10(1): 1-6.
- [3] 左军成, 陈宗镛, 周天华. 中国沿岸相对海面变化的本征分析和预测[J]. 海洋与湖沼, 1995, 26(3): 331-337.
- [4] 郑大伟, 虞南华. 上海地区海平面上升趋势的长期预测研究[J]. 中国科学院上海天文台年刊, 1996, 17: 36-45.
- [5] 秦曾灏, 李永平. 上海海平面变化规律及长期预测方法的初探[J]. 海洋学报, 1997, 19(1): 1-7.
- [6] 左军成, 陈宗镛, 戚建华. 太平洋海域海平面变化的灰色系统分析[J]. 青岛海洋大学学报, 1997, 27(2): 138-144.
- [7] 夏华永, 李树华. 带周期项的海平面变化灰色分析模型及广西海平面变化分析[J]. 海洋学报, 1999, 21(2): 9-17.
- [8] Daubechies I. Ten Lectures on Wavelets[M]. Philadelphia: CBMS NFS Series in Applied Mathematics, SIAM, 1992. 129-214.
- [9] Mallat S. A theory for multiresolution signal decomposition: The wavelet representation[J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 1989, 11(7): 674-693.
- [10] Muzzy J F, Barcery E, Meodo A A. The multifractal formalism revisited with wavelets[J]. International Journal of Bifurcation and Chaos, 1994, 4: 245-302.
- [11] 冯文权. 经济预测与决策技术[M]. 武汉: 武汉大学出版社, 1996. 295-297.

(下转第 51 页)

Prediction experiment of sea-level changes based on the wavelet analysis and auto-regressive model

YUAN Lin wang, XIE Zhi ren, ZHONG He-xiang

(Geor Science School of Nanjing Normal University, Nanjing 210097, China)

Received: Nov. , 24, 2004

Key words: sea level change; prediction; wavelet analysis(WA); auto regressive(AR) model

Abstract: Based on the monthly average tidal records of Wusong tidal gauge station from 1955 to 2001, a prediction experiment is made using wavelet analysis (WA) and auto-regressive(AR) models in this paper. WA provides a way of analyzing the local behavior of functions, it may be used to isolate the sea-level change records to reveal the hierarchy of features and the scaling behavior. Since components are the filtered versions of the raw data, their behavior is more regular than that of the original signal and more predictable accordingly. Using AR model to predict each component in practice we can take advantage of each significant component, the summation of all the predicted components is the predict results. Analyzing the data of 1955-1996 to predict the sea level change of 1997-2001, comparing the prediction results with the original data, we find that these two series are fairly comparable. It was shown that the integration use of these two methods provides an efficient way to predict sea-level changes.

(本文编辑:刘珊珊)