

白冠统计物理模型的验证

——I. 二维等间距投影图的数字化方法研制与应用

郁诚成^{1,2,3}, 孙宝楠^{1,2,3*}, 杨永增^{1,2,3}, 连展^{1,2,3}

(1. 国家海洋局 第一海洋研究所, 山东 青岛 266061;

2. 青岛海洋科学与技术国家实验室 区域海洋动力学与数值模拟功能实验室, 山东 青岛 266237;

3. 海洋环境与数值模拟国家海洋局重点实验室, 山东 青岛 266061)

摘要:近几十年以来,随着海洋观测技术的提高,海洋观测数据大量增加,极大地推动了海洋研究的发展,但由于某些原因,部分数据只以图像形式对外发布,很难获取其原始数据,这给科研工作的及时开展带来了诸多不便。本文基于 MATLAB 软件的图像和数据处理功能,提出了一种二维等间距投影图像的数字化方法,可实现图像信息快速准确的抓取,并以 Levitus 的全球海表温度为例,验证了数字化过程的可靠性。数字化结果与原始数据比较显示,平均误差为 0.147 °C,绝对平均误差为 0.215 °C,均方根误差为 0.275 °C,数字化结果与原始数据吻合较好。最后,本文将该方法应用到 2001-02,2001-05,2001-08 和 2001-11 卫星反演的白冠覆盖率图像中,所获数字化结果将应用到下一步统计物理白冠模型的验证工作中。

关键词:图像处理; SST 图像数字化和检验; 白冠图像数字化; MATLAB 软件

中图分类号: p731.22

文献标识码: A

文章编号: 1671-6647(2017)03-0362-07

doi: 10.3969/j.issn.1671-6647.2017.03.006

近几十年以来,海洋观测技术发展迅速,特别是随着卫星观测技术的成熟,海洋数据大量增加,极大地推动了海洋研究的发展。很多类型的海洋数据可以通过网络或其他方式直接获取,但同时,还有一部分数据只以图像的形式向外发布,很难直接获得原始数据,比如不公开算法的卫星反演图、人工绘制的历史图等。此类图像大都具有非常重要的研究价值,但因其数据不易获取给科研工作的开展带来了极大的不便。随着计算机技术的发展,人们发现图像数字化方法可以解决这一问题。图像数字化是将连续色调的模拟图像经采样量化后转换成数字影像的过程。目前,常见的点、线图形的数字化应用较为广泛,但对于二维图像没有直接的工具可以对其数据进行抓取,相关文献也鲜有报道,所以研究一种二维图像的数字化方法并在将来形成一个可方便研究人员直接调用的程序包,就显得格外有意义。

在以往的研究中,人们常用 MATLAB 软件进行图像的研究,因其具有图像处理功能强大、功能全面、界面友好、开放性等特点,不仅有助于提高研究效率,还有助于加快科研进展^[1-3]。邓巍等^[4]利用 MATLAB 图像工具箱实现了不同类型的图像的处理;周广芬等^[5]利用 MATLAB 图像处理工具箱对一维数据图像进行了数据化;郭磊等^[6]结合 MATLAB 图像数值分析程序和专业位图软件 PHOTOSHOP 图像操作模块,实现了对一维潮位图的数字化处理;崔红等^[7]利用 MATLAB 对尾流图像进行了灰度变换、滤波等数字化处

收稿日期: 2016-08-25

资助项目: 国家高技术研究发展计划项目——南海与周边海域风浪流耦合同化精细化数值预报与信息服务系统(2013AA09A506); 国家自然科学基金委员会-山东省人民政府联合资助海洋科学研究中心项目——海洋环境动力学和数值模拟(U1406404); 山东省自然科学基金项目——南海内波致跨等密度面湍流混合的数值模拟研究(2015ASKJ01); 国家自然科学基金青年基金项目——南海北部温跃层季节内变化特征及其与海洋涡旋的关系研究(41506037); 国家海洋局中国-印尼海上合作基金子课题-青年基金项目——南海及其周边海域内波致混合数值模拟研究(QY0516015)

作者简介: 郁诚成(1992-),男,浙江宁波人,硕士研究生,主要从事物理海洋学方面研究. E-mail: yucc@fio.org.cn

* **通讯作者:** 孙宝楠(1984-),女,山东青岛人,助理研究员,在职博士,主要从事物理海洋学方面研究. E-mail: sunbn@fio.org.cn

(李 燕 编辑)

理,为尾流图像进入信息化系统提供了基础;夏长水和袁业立^[8]对塘沽地区的 SAR 图像进行了数字化,并应用到地形的反演研究中;王淑娟和张玉叶^[9]以 MODAS 的 SST 数据为例,介绍了一种从索引图像中提取数据的方法。宋玮等^[10]通过像素值对南沙双子裙礁 SAR 和 TM 图像进行了噪声滤除和几何校正处理。上述大部分研究中,或只对 MATLAB 在图像处理中的功能做了介绍,没有运用到图像的数字化中;或只对一维数据图像做了数字化处理和检验;鲜有实现将二维图像转化为实际数据的完整的方法讨论、数字化展示以及相应的结果验证。究其原因,二维图像的量化过程(即色标颜色值与数据值的对应关系以及颜色值到数据值的转换环节)较为复杂。同时二维图像的投影种类繁多,按照投影的变形性质可以分为等角投影、等积投影和任意投影。其中等距投影是任意投影中比较特殊的投影,存在正轴投影中经线长度比为 1,在斜轴或横轴投影中垂直圈长度比为 1 的性质。本文将数字化方法应用于等间距投影的二维图像,给出一个较为实用的二维等间距投影图的数据获取方法,并验证数字化结果的可靠性,以期为此类问题提供一个有效的解决方案。

1 图片数字化方法研究

1.1 数据介绍

Levitus^[11-12]是由美国国家海洋数据中心(NODC)的海洋气候实验室制作的气候态海洋数据,是目前国际上比较常用的数据资料。具有年平均、季节平均和月平均三种不同时间尺度,空间分辨率为 1° ,范围覆盖全球,垂直方向分为 33 层。本文实验中使用的是 1 月月平均的全球海表温度资料(图 1)。

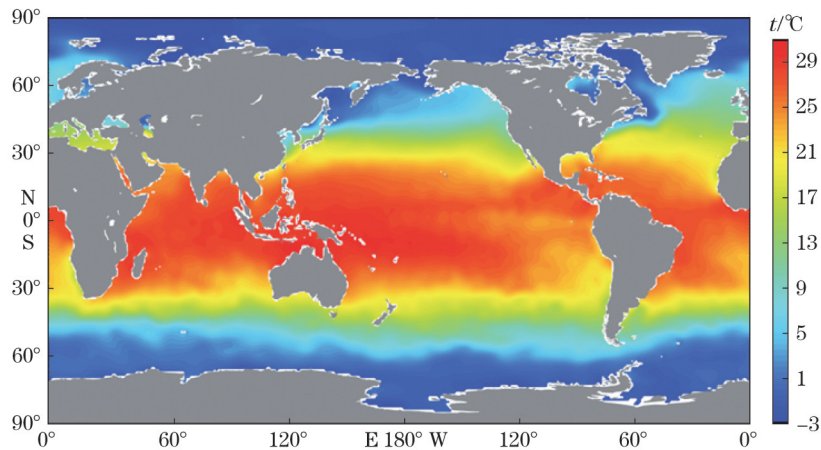


图 1 Levitus 全球 1 月月平均海表温度图

Fig.1 Levitus monthly averaged sea surface temperature (SST) of January

1.2 方法介绍

MATLAB 可以读取图像的 RGB 值,如果要获取 RGB 对应的数据值,应将 RGB 颜色值与数据值的关系确定出来,通常这些关系包含在色标的内容中。本部分内容以 Levitus 的全球海表温度为例,先将 Levitus 原始数据图像化,得到 RGB 图像,再从 RGB 颜色值与温度数据值中找到一一对应关系,重现图像的数据。将重现数据与原始数据做了误差分析,得到了可靠的图像信息。具体过程如下:

1) 建立像素值与颜色值的对应关系

欲将图像转化为数据,需要知道每种颜色所对应的数据值的大小。所以,首先利用 MATLAB 将色标图像读入,以色标外边框为界对图像进行剔除,选取有效区域 RGB 数据,依次提取各个颜色值,再根据色标的

下标确定每种颜色对应的数据值,得到颜色值和数据值的一一对应关系。以上处理过程函数为

$$\text{function cb}=\text{loc_cb}(\text{Filename},\text{val_init},\text{val_delt}). \quad (1)$$

输入参数 *Filename* 为 colorbar 图像名, *val_init* 为 colorbar 第一个颜色所对应的数据值, *val_delt* 为 colorbar 值的间隔。

输出量 *cb* 是一个 cell 数组 $\text{cb}\{[\text{R G B}], \text{value}\}$, $\text{cb}\{N,1\}$ 为颜色值, $\text{cb}\{N,2\}$ 为对应的数据值。通过该功能函数建立了颜色值与数据值之间的一一对应关系。

2) 颜色值到数据值的转换

确定颜色值和数据值的关系后,为将图像中的颜色值转换到数据值做好了准备。首先读入原始图像,并将图像的边框及以外部分剔除,即只保留含有数据的最小矩形区域。值得注意的是,由于海洋和大气数据图像往往存在陆地区域,因此需从原始图像中截取陆地区域图像,同样通过 MATLAB 来读取代表陆地的颜色值,或直接给定输入陆地的颜色值,这样在数字化的过程中可以剔除陆地的像素点,只对海水点进行处理。

完成数据区域选取的处理,之后,就可以将原始图像中含有数据的像素点,通过颜色值和数据值的一一对应关系,将原始图片的颜色值转换为数据值。有时因图像本身原因或图像采集过程,某些点的 RGB 值发生些许变化,无法与原色标中的 RGB 值相匹配,产生转换不成功现象。为此采用 CRESSMEN 插值^[13]进行转换失败点的数据补全。同时,若原始图像中有等值线标注,等值线的颜色值不同于陆地点的颜色值又无法与色标中的颜色值匹配,等值线所在的像素点也会以插值的形式补全。最后,由于本次方法应用于等间距投影图,将给定的坐标范围均匀的分配到每个像素点上,就完成了像素点的坐标确定。以上处理过程函数为

$$[\text{result},\text{mask},\text{lon},\text{lat}]=\text{img2data}(\text{Picname},\text{cb},\text{lon_range},\text{lat_range},\text{varargin}). \quad (2)$$

输入参数 *Picname* 为待提取数据图像的图像名; *cb* 为上一函数的输出量,即颜色值与数据值的对应关系; *lon_range* 和 *lat_range* 分别为数据的经纬度范围,如: $\text{lon_range}=[90\ 270]$, $\text{lat_range}=[-45\ 45]$,表示经度范围为 90° 到 270° ,纬度范围为 -45° 到 45° ,可选输入量 *varargin* 可以是陆地图像或陆地颜色的 RGB 值。

输出量包括 *result*, *mask*, *lon* 和 *lat*。Result 为 $M \times N$ 的数组, *M* 和 *N* 分别为图像的纵横像素点数的个数,储存各个点的数据值,若值为 32 767,则该点为陆地点或缺测点,即无数据点; *mask* 为 $M \times N$ 个像素点的水陆判别信息,1 代表水点,0 代表陆地点; $\text{lon}(N)$ 为横向各个像素点的经度值, $\text{lat}(M)$ 为纵向各个像素点的纬度值。

图 2 为数字化数据结果的再现图,图 3 为数字化方法的流程图。

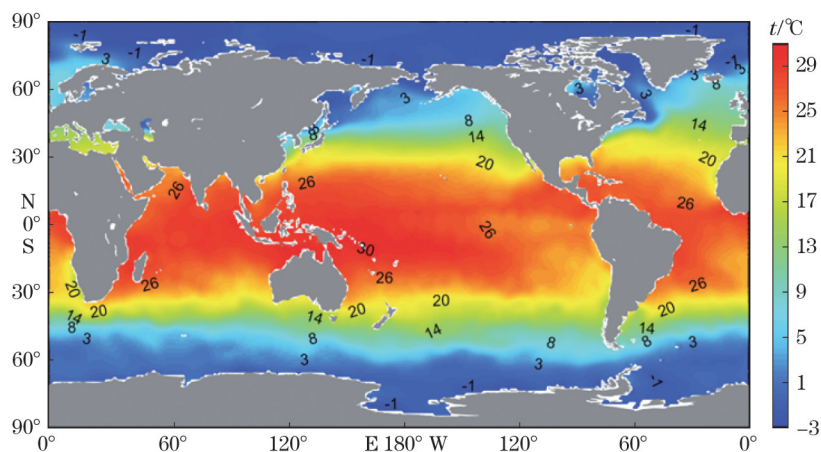


图 2 数字化结果再现图

Fig.2 Map based on digitized SST

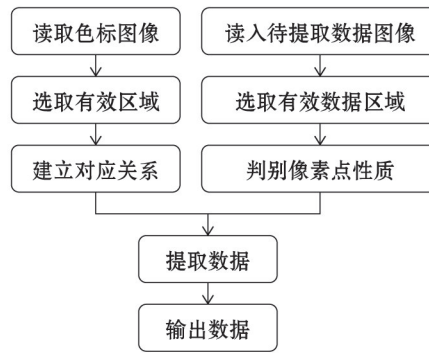


图3 数字化方法流程图

Fig.3 Flow chart of the digitization method

1.3 图像数字化模型检验

图4为原始数据与数字化结果的散点对比图。可以看出数字化的提取结果与原始数据具有较好的一致性。图中极个别点偏差较大,且相对于原始数据都偏小。原因在于这些点为水陆交界点,并且在数据提取中匹配失败,是由插值补全得到,而由于陆地点的影响使所得数据变小。经计算,原始数据与再现数据平均误差(ME)为 $0.147\text{ }^{\circ}\text{C}$,绝对平均误差(MAE)为 $0.215\text{ }^{\circ}\text{C}$,均方根误差(RMSE)为 $0.275\text{ }^{\circ}\text{C}$,相关系数(R)为 $0.999\ 86$,结果良好。故总体而言,通过本文的数字化方法提取的数据是可信的,可以运用到进一步的研究工作中。

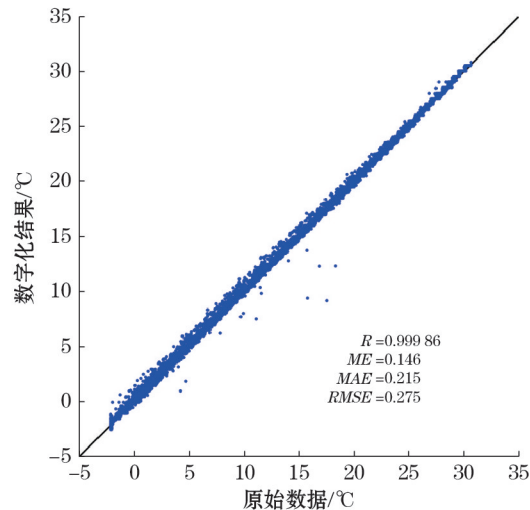


图4 Levitus 原始数据与数字化结果的散点对比图

Fig.4 A scatter plot of Levitus SST and digitized SST

2 图片数字化方法实际应用

将该方法应用到全球月平均白冠覆盖率图像^[14],此处只给出2001-02,2001-05,2001-08和2001-11(图5a~5d)四个月份的原始图像并给出相应月份的数据化结果再现图像(图5e~5h)。图像原始数据是从卫星资料反演得到,其中包括SSM/I的海表面亮温(T_B)、大气柱水汽含量(V)、云中液态水含量(L)数据、AVHRR的海表面温度(T_s)数据、WOA05的海表面盐度(S)数据和QuikSCAT的海面10 m风场(U_{10})数据等。该反演方法较复杂,对于物理模型的验证来说非常耗时。

按照上部分图片的数字化方法,经过选取色标颜色值,建立颜色值与数据值对应关系,选取数据矩形区域,剔除陆地点,颜色值与数据值的转换,经纬度的确定等一系列处理,最终的数据提取结果如图5e~5h所示。从图中可以发现,数字化再现图像与原图像的空间布局基本一致,数据大小基本吻合,提取结果较为满意,实现了快速获取图像信息的功能。

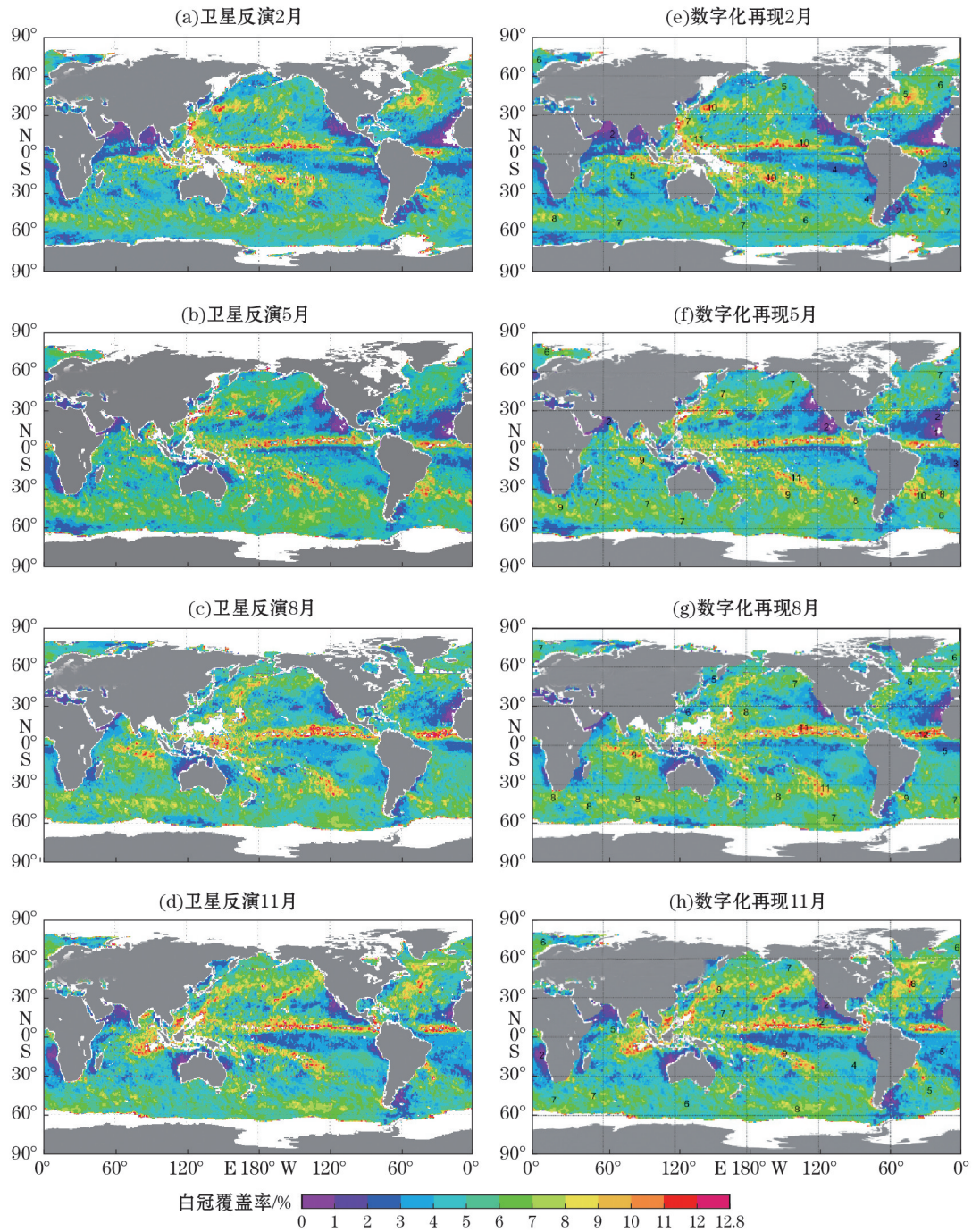


图5 由卫星数据反演得到(a~d)^[14]和数字化再现(e~h)的2001-02,2001-05,2001-08和2001-11月平均全球白冠覆盖率图像

Fig.5 Monthly average global whitecap coverage images in February, May, August, and November 2001 based on satellite measurements(a~d)^[14] and digitized data(e~h)

3 结 语

本文的数字化方法能够一站式的对图像进行有效的数据提取,将图像的像素值转化为数据值。该数字

化方法应用于等间距投影的二维图像,并且主要针对因某些原因只向外提供图像但没有原始数据的情况,如原始数据具有较高研究价值的卫星图像,绘制时间较早且原始数据丢失的图像,人工绘制的历史图像等。该方法数字化过程包括色标的确定、待提取数据图像的处理、数据提取和坐标确定等几个步骤。

上述图像数字化方法可为此类无法直接获取的数据提供便捷的数字化,相对于图像,数字化后得到的数据更具有灵活性。并且,若进行数字化的图像包含的数据范围较大,如全球范围的数据,其原始数据的分辨率一般较低,而数字化过程中将数据范围均匀分配到各个像素点上,对于一般分辨率的图像,横纵像素点个数相对较大,因此最后提取得到的数据分辨率相对与原始数据有所提高,而对于包含较小数据区域的图像,原始数据分辨率相对较高,通过数字化提取的数据也能保持相当的分辨率。

同时本文对数字化提取结果的有效性进行了检验,原始数据与再现数据平均误差为 $0.147\text{ }^{\circ}\text{C}$,绝对平均误差为 $0.215\text{ }^{\circ}\text{C}$,均方根误差为 $0.275\text{ }^{\circ}\text{C}$,相关系数为 $0.999\ 86$,检验结果较为满意,可将数据运用到进一步的研究工作中以供后续分析使用。

目前上述提出的方法仅仅适用于空间等间距投影图像,对于那些空间非等间距投影的图像暂时无法取得令人满意的结果,有必要对该类问题做进一步的研究。

参考文献(References):

- [1] XU Z Y, LI J P. MATLAB and its application to digital image processing[J]. Computer and Modernization, 2003(4): 64-65,69. 许志影,李晋平. MATLAB 及其在图像处理中的应用[J]. 计算机与现代化, 2003(4): 64-65,69.
- [2] TU W M, WEI Y G, SHI S M. Application of MATLAB to digital image processing[J]. Microcomputer Information, 2007, 23(2X): 299-300. 涂望明,魏友国,施少敏. MATLAB 在数字图像处理中的应用[J]. 微计算机信息, 2007, 23(2X): 299-300.
- [3] CHU Q Y. The application of MATIAB to image process[J]. Information Technology & Informatization, 2010(4): 55-56. 初青瑜. MATLAB 在图像处理中的应用[J]. 信息技术与信息化, 2010(4): 55-56.
- [4] DENG W, DING W M, ZHANG H. Application of MATLAB in figure image processing and analysis[J]. Journal of Agricultural Mechanization Research, 2006(6): 194-198. 邓巍,丁为民,张浩. MATLAB 在图像处理和图分析中的应用[J]. 农机化研究, 2006(6): 194-198.
- [5] ZHOU G F, LI P, YANG J Y. Study of using the image processing toolbox of MATLAB to digitize graph[J]. Journal of Hebei University of Science and Technology, 2005, 26(4): 309-311,321. 周广芬,李鹏,杨久义. 利用 MATLAB 图像处理工具箱进行图形数字化的研究[J]. 河北科技大学学报, 2005, 26(4): 309-311,321.
- [6] GUO L, HUANG F H, QIU J. Method and application of tide level digitizing by MATLAB image processing[J]. Journal of Yangtze River Scientific Research Institute, 2014, 31(7): 86-90. 郭磊,黄锋华,邱静. 基于 MATLAB 图像处理的潮位数字化方法及应用[J]. 长江科学院院报, 2014, 31(7): 86-90.
- [7] CUI H, CHANG Y, ZHANG J S. Digital processing of wake images based on MATLAB[J]. Acta Photonica Sinica, 2010, 39(12): 2274-2278. 崔红,常洋,张建生. 基于 MATLAB 的尾流图像数字化处理[J]. 光子学报, 2010, 39(12): 2274-2278.
- [8] XIA C S, YUAN Y L. Simulation and inversion study on the submarine topography in the Tanggu sea area using SAR images[J]. Advances in Marine Science, 2003, 21(4): 437-445. 夏长水,袁业立. 塘沽海区海底地形的 SAR 影像仿真与反演研究[J]. 海洋科学进展, 2003, 21(4): 437-445.
- [9] WANG S J, ZHAGN Y Y. Extracted data from image by MATLAB[J]. Science & Technology Information, 2010(11): 49-50. 王淑娟,张玉叶. 利用 MATLAB 从图像中提取数据[J]. 科技信息, 2010(11): 49-50.
- [10] SONG W, ZHANG J, JI G R, et al. Characteristic analysis of Nansha Shuangzi Reefs based on SAR and TM images[J]. Advance in Marine Science, 2004, 22(Supp. 1): 177-181. 宋玮,张杰,姬光荣,等. 基于 SAR 与 TM 图像的南沙双子群礁特征分析[J]. 海洋科学进展, 2004, 22(增 1): 177-181.
- [11] DOU C E, LU J K, TIAN H J. Analysis of the seasonal characteristics of the upper layer water quality at South Yellow Sea based on LEVITUS94[J]. Journal of Huaihai Institute of Technology(Natural Science), 2010, 19(3): 84-88. 窦长娥,路吉坤,田慧娟. 基于 LEVITUS94 资料的南黄海上层水质季节变化特征分析[J]. 淮海工学院学报(自然科学版), 2010, 19(3): 84-88.
- [12] LEVITUS B S. World Ocean Atlas Vol 1-5[C]//NOAA Atlas NESDIS. Washington D C: U. S. Department of Commerce, 2010.
- [13] CRESSMAN G P. An operational objective analysis system[J]. Monthly Weather Review, 1959, 87(10): 367-374.
- [14] YU T. The Preliminary study of whitecap coverage estimated from routine satellite measurements over the global[D]. Qingdao: Ocean University of China, 2008: 47-50. 于婷. 由卫星资料反演全球白冠覆盖率的初步研究[D]. 青岛: 中国海洋大学, 2008: 47-50.

Validation of Statistical Physics Models for Whitecap:

I. Development and Application of Digitization Technique of 2D Equidistant Projection Image

YU Cheng-cheng^{1,2,3}, SUN Bao-nan^{1,2,3}, YANG Yong-zeng^{1,2,3}, LIAN Zhan^{1,2,3}

(1. *The First Institute of Oceanography, SOA, Qingdao 266061, China;*

2. *Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine
Science and Technology, Qingdao 266237, China;*

3. *Key Laboratory of Marine Science and Numerical Modeling, SOA, Qingdao 266061, China)*

Abstract: In recent decades, with the rapid development of ocean observing technology, marine data has increased rapidly and given a big boost to the marine research. However, some of those data were published only in the form of images, which are difficult to get the original data, bringing a lot of inconvenience to research work. Basing on the image and data processing function of MATLAB software, we developed a digitization method for 2D equal-distant projection image. Utilize the global sea surface temperature (SST) from Levitus as an example, reliability of our digitization method is verified. Compared with the original SST, the average error of the digitized SST is 0.147 °C, the averaged absolute error is 0.215 °C, and the root mean square error is 0.275 °C. The digitized SSTs agree well with the original data, and thus can be used for further study. Finally, the digitization method was applied to whitecap coverage in satellite image of February, May, August, and November 2001, and the digitized results will be further used in verification of a statistical physics model for whitecap simulation.

Key words: image processing; SST image digitization; whitecap image digitization; MATLAB software

Received: August 25, 2016