doi:10.3969/j.issn.1003-2029.2019.04.002

总体最小二乘法在浮游植物色素吸收系数模型估算中的应用

李芝凤,李铜基,朱建华
(国家海洋技术中心,天津 300112)

摘 要:采用常规最小二乘法(LS)和总体最小二乘法(TLS)对浮游植物色素吸收系数与叶绿素 a 浓度之间的关系进行了研究。利用 2003-2012 年 6 个航次 243 组实测数据,建立了 LS 和 TLS 模型,应用模型估算典型波段的色素吸收系数值,并对两种模型进行了验证。采用平均绝对百分误差和均方根误差两种方法进行模型效果评估,结果显示,两种评估方法均能很好地说明,在 400~700 nm 光谱范围内,TLS 拟合结果优于 LS 拟合结果。总体最小二乘法在综合考虑了参数误差和测量误差的情况下,拟合结果精度更高,更符合实际情况,估算结果的准确性要优于常规最小二乘法。

关键词:色素吸收系数;叶绿素 a 浓度;最小二乘法;模型;评估
中图分类号:TP79 **文献标志码**:A **文章编号**:1003-2029(2019)04-0009-06

浮游植物色素吸收系数是水体的固有光学参数之一,是描述浮游植物生长和初级生产力的基本参数,体现了浮游植物的光吸收特性。叶绿素 a 是浮游植物的最主要色素,其浓度与吸收系数之间有着密切的联系。研究二者之间的关系,是水色遥感生物-光学算法开发的重要组成部分[1-2]。受区域、季节、种群、色素组成和包囊效应等多种因素的影响,浮游植物色素吸收系数与叶绿素 a 浓度的关系非常复杂,多为非线性的,国内外学者将二者的关系描述为幂指数关系,如式(1)所示。

$$a_{ph}(\lambda)=A(\lambda)[Chla]^{E(\lambda)} \tag{1}$$

式中:$a_{ph}(\lambda)$为浮游植物色素吸收系数;$[Chla]$为叶绿素 a 浓度;参数 $A(\lambda)$反映了 $a_{ph}(\lambda)$的数值大小;而指数 $E(\lambda)$则反映了$[Chla]$的变化对 $a_{ph}(\lambda)$的响应关系,也可以从某种程度上反映出浮游植物包囊效应的强弱[2]。通过实测 $a_{ph}(\lambda)$和$[Chla]$数据建立模型,就可获得参数 $A(\lambda)$和 $E(\lambda)$,进而得出 $a_{ph}(\lambda)$和$[Chla]$之间的关系。

Bricaud 等[3-4]对马尾藻海、地中海、河口与海湾、大西洋等海区的一类和二类海水进行了综合分析,结果表明,色素吸收系数和叶绿素 a 浓度之间存在很好的幂指数关系;Matsuoka A 等[5]对北极西部海域浮游植物、NAP 和 CDOM 的光吸收系数的变化趋势进行了研究,采用幂函数来检验各组分的吸收系数与叶绿素 a 浓度之间的关系,同样得到了很好的效果;Churilova T 等[6]利用幂指数的最小二乘拟合函数推导出黑海水域不同季节色素吸收系数与叶绿素 a 浓度之间的经验关系,并将其应用到卫星数据对叶绿素 a 浓度的估算中;Meler J 等[7]利用色素吸收系数与叶绿素 a 浓度之间的幂指数关系,分析了波罗的海南部和中部不同海区、不同季节的参数变化。在国内,朱建华等[1]利用 2003 年黄东海春、秋两个航次色素吸收系数和叶绿素 a 浓度的数据研究了二者之间的关系,结果表明,二者之间存在幂指数关系, 并利用叶绿素 a 浓度反演色素吸收系数,效果良好;曹文熙等[8]在对珠江口悬浮颗粒物的吸收光谱及区域模式的研究中,将浮游植物色素吸收与叶绿素 a 浓度的关系描述为幂指数关系,吸收系数明显地随叶绿素 a 浓度呈非线性增加;王林等[2]

收稿日期:2019-05-27
作者简介:李芝凤(1978-),女,工程师,主要从事海洋水体光学特性研究。E-mail:lzfeng.tj@126.com

利用 2007 年北黄海海域的现场实测数据研究了吸收系数与叶绿素 a 浓度之间的关系，结果表明，675 nm 波段为参考波段时乘幂拟合相关性明显高于 440 nm 波段。

目前，对色素吸收系数与叶绿素 a 浓度之间幂指数关系的拟合，采用的多是常规最小二乘法（Least Squares，LS），总体最小二乘法（Total Least Squares，TLS）的应用以及两种方法拟合结果的比较尚未发现。本文利用 2003–2012 年间 6 个航次测得的浮游植物色素吸收系数和叶绿素 a 浓度的现场数据，选择 412，443，490，510，555，675 nm 为参考波长，采用常规最小二乘法和总体最小二乘法进行拟合，建立浮游植物色素吸收系数的 LS 和 TLS 估算模型，将实测值与模型估算结果进行了比对，结果显示 TLS 拟合效果优于 LS 拟合效果。

1 研究区域

本文研究区域选择在黄东海，水样分别采集于 2003 年 4 月、2003 年 9 月、2006 年 7 月、2006 年 12 月、2011 年 5 月和 2012 年 4 月，共 6 个航次 243 个样品，站位分布如图 1 所示。

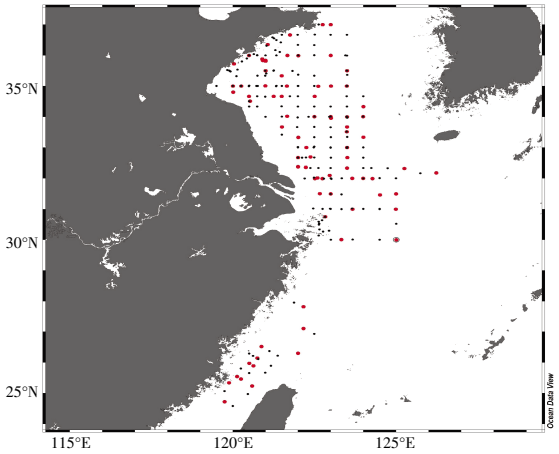


图 1 站位分布图

2 研究方法与数据获取

2.1 研究方法

总体最小二乘法是一种能同时兼顾因变量和自变量中随机误差的参数估算方法，既可以对观测向量进行误差改正，也可以对自变量组成的系数矩阵进行误差改正，能大大减弱随机误差对参数估计造成的影响[9]。应用 TLS 法进行拟合，自 Golub 于 20 世纪 80 年代提出以来，经过几十年的发展，已广泛应用于统计分析、线性和非线性回归、系统识别和参数估计等相关领域[10]。虽然常规最小二乘法是目前对观测数据进行处理最基本、最常用的方法，但是应用的前提是自变量为精确值，计算过程忽略了自变量的误差。由于浮游植物色素浓度与其光谱吸收系数为现场测量数据，均存在测量误差，尤其当测量误差较大时，经典最小二乘法并不能给出符合实际情况的拟合结果。

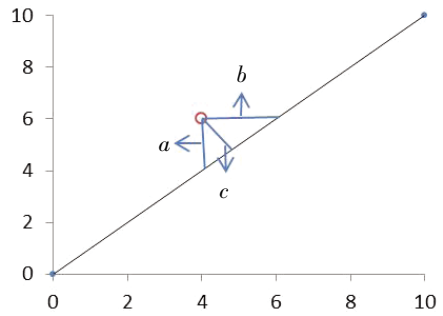
下面以一元线性回归为例说明总体最小二乘法的基本思想。

$$y_i = ax_i + b \tag{2}$$

式中：a 为理论回归系数；x_i 为自变量；y_i 为因变量，(i=1，2，…，n)。考虑到自变量 x_i 和因变量 y_i 同时存在测量误差，则式(2)可表示为：

$$y_i + v_{y_i} = a\,(x_i + v_{x_i}) + b \tag{3}$$

式(3)即为总体最小二乘法的基本模型，v_{x_i} 和 v_{y_i} 分别为 x_i 和 y_i 产生的随机误差。若 x_i 能够准确测量，则 $v_{x_i}=0$，测量误差仅存在于变量 y_i 中，因此，可用经典最小二乘法求解 y_i；若 y_i 能够准确测量，则同样可用经典最小二乘法求解 x_i，此时测量值与计算值之间的误差平方和最小。而在很多数据处理的过程中，自变量和因变量同时存在测量误差，即 $v_{x_i}\neq 0$，$v_{y_i}\neq 0$，参数估计值 a 可通过使各实验点到拟合直线的垂直距离的平方和 $\Sigma(y_i-ax_i)^2/(1+ax_i)^2$ 最小来求得。LS 和 TLS 的几何解释如图 2 所示。



a：$v_{x_i}=0$(LS)；b：$v_{y_i}=0$(LS)；c：$v_{x_i}\neq 0$，$v_{y_i}\neq 0$(TLS)

图 2 LS 和 TLS 模型的几何解释

2.2 数据获取

每个站位用采水器采集表层两份平行水样，一份用于测定浮游植物色素吸收系数，另一份用于测定叶绿素 a 浓度。浮游植物色素吸收系数利用定量滤膜分光光度法[11]获得：将水样过滤到 GF/F 玻璃纤

维滤纸上,获得总颗粒物样品滤纸,利用分光光度计(GBC Cintra20)先后测量总颗粒物和甲醇提取后的非色素颗粒物的光学密度,计算出总颗粒物和非色素颗粒物的吸收系数,二者之差即为浮游植物色素吸收系数 $a_{ph}(\lambda)$。叶绿素 a 浓度的测量,将现场水样过滤至 GF/F 滤纸上,用 90%丙酮提取色素,利用 HPLC(Waters 600E)对提取液中的叶绿素 a 进行分离和定量分析[8]。

3　结果与讨论

3.1　不同航次的浮游植物色素吸收系数与叶绿素 a 浓度

表 1 和图 3 分别列出了 6 个航次 243 个站位表层叶绿素 a 浓度的数据统计，浓度范围为 0.11~8.03 μg/L,164 个站位的叶绿素 a 浓度小于 1.0 μg/L,占总站位数的 67.5%,2003 年 4 月和 2003 年 9 月航次，叶绿素 a 浓度较其他航次数值偏高,2003 年 9 月航次出现最高值,达到 8.03 μg/L;浮游植物色素吸收系数光谱(图 4)在 440 nm 和 675 nm 附近出现吸收峰值，吸收系数数值一般都在 0.2 m^{-1} 以下,2003 年 4 月航次部分站位色素吸收系数高于其他航次,变化范围为 0.2~0.7 m^{-1}。

表 1　航次信息与叶绿素 a 浓度

航次	时间	最小值 /μg·L⁻¹	最大值 /μg·L⁻¹	平均值 /μg·L⁻¹	数据量
200304HDH	2003 年 4 月	0.16	6.83	1.66	51
200309HDH	2003 年 9 月	0.16	8.03	1.66	48
ST03X	2006 年 7 月	0.11	3.44	0.76	128
ST05D	2006 年 12 月	0.44	0.83	0.60	6
201105CJK	2011 年 5 月	0.19	0.73	0.49	6
201204CJK	2012 年 4 月	0.20	1.63	0.80	4

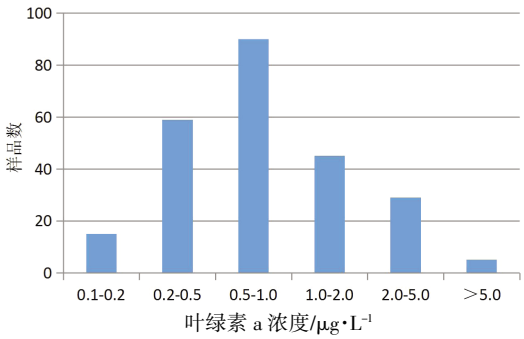


图 3　叶绿素 a 浓度数据统计

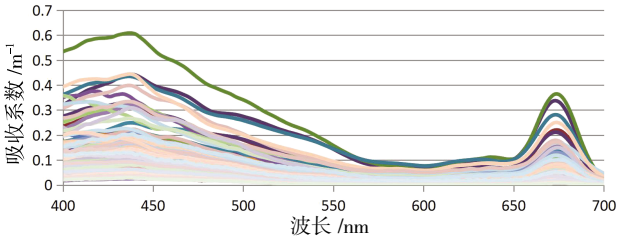


图 4　浮游植物色素吸收系数

本文共采集数据 243 组,随机选取 162 组数据(约占总数据量的 2/3)进行 LS 和 TLS 模型开发,利用其余 81 组数据对模型进行验证和效果评估。模型开发数据和模型验证数据的站位分布如图 1 所示,图中蓝色为模型开发站位,红色为模型验证站位。

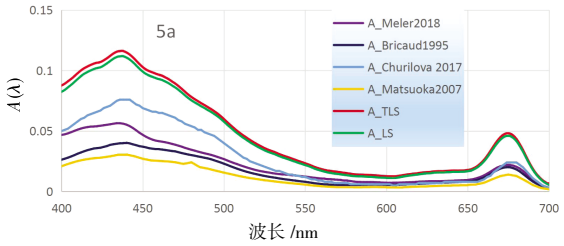
3.2　最小二乘法拟合

采用 LS 和 TLS 方法,逐波长计算出 $A(\lambda)$和 $E(\lambda)$(表 2),并在 400~700 nm 波长范围内作图,如图 5 所示,图 5(a)为参数 $A(\lambda)$结果,图 5(b)为指数 $E(\lambda)$结果。从图中可以看出,参数 $A(\lambda)$的波形与吸收系数 $a_{ph}(\lambda)$(图 4)的波形一致,在 440 nm 和 675 nm 附近出现峰值。LS 和 TLS 拟合曲线在 400~700 nm 波长范围内变化趋势一致,A_TLS 数值稍高于 A_LS;将 A_LS 和 A_TLS 的拟合结果与文献[3,5-7]的研究结果比较,本文结果与 Churilova 等[6]的研究结果最为相似,波形变化趋势一致,但数值偏大,这可能与季节、研究海域的生物种群及包囊效应有关。

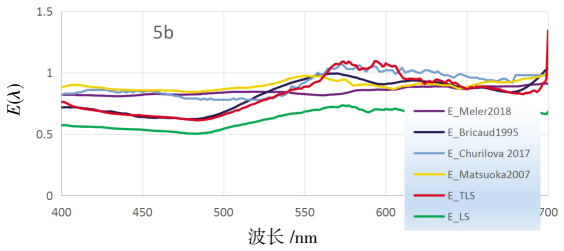
指数 $E(\lambda)$反映了叶绿素 a 浓度的变化对吸收系数的响应关系,从图 5(b)中可以看出,E_TLS 数值明显大于 E_LS,这是由于 LS 和 TLS 计算方法不同

表 2　参数 $A(\lambda)$和指数 $E(\lambda)$计算结果

波长/nm	LS		TLS	
	$A(\lambda)$	$E(\lambda)$	$A(\lambda)$	$E(\lambda)$
412	0.095 418	0.561 012	0.100 075	0.709 589
443	0.108 028	0.539 578	0.112 100	0.654 958
490	0.067 434	0.512 640	0.069 853	0.622 532
510	0.046 297	0.573 399	0.048 217	0.700 102
555	0.019 331	0.696 466	0.020 714	0.912 026
675	0.046 088	0.701 827	0.048 231	0.843 582



(a) 参数 $A(\lambda)$结果



(b) 指数 $E(\lambda)$结果

图 5 LS、TLS 拟合 $A(\lambda)$和 $E(\lambda)$以及与国外部分研究结果的比对

导致的。TLS 方法拟合,同时考虑了参数误差和自变量即叶绿素 a 浓度的测量误差,因此,得出的结果更加符合实际情况。本文结果与 Churilova 等[6]的研究结果变化趋势一致,400~520 nm 之间,E_TLS 数值偏低,520~700 nm 之间数值基本一致。

根据浮游植物色素吸收系数和叶绿素 a 浓度实测值,对 412,443,490,510,555,675 nm 等典型波长进行拟合,效果如图 6 所示。图中绿色虚线为

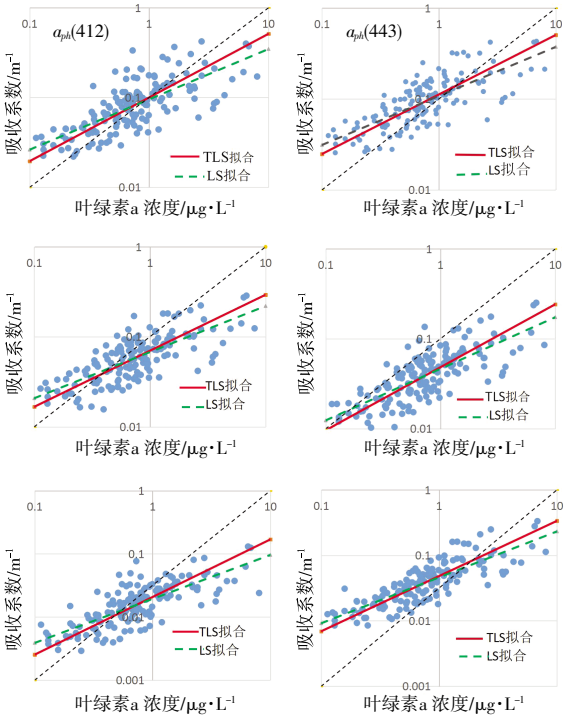
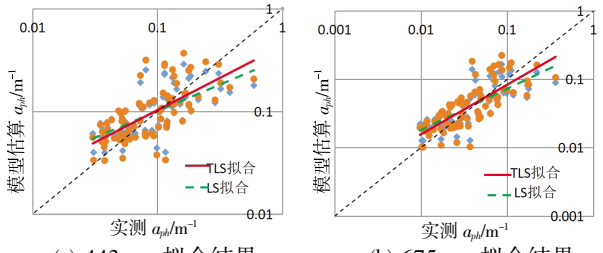


图 6 典型波长吸收系数与叶绿素 a 浓度的 LS、TLS 拟合效果

LS 拟合结果,红色实线为 TLS 拟合结果。从图中可以看出,TLS 的拟合效果优于 LS 拟合效果,412 nm、443 nm 和 675 nm 拟合效果优于其他波段。LS 拟合时,忽略了叶绿素 a 浓度的测量误差,而叶绿素 a 浓度,从采样、过滤、样品前处理到仪器测量的过程中均存在误差,并且叶绿素 a 浓度测量结果的变化范围很大(在 0.11~8.03 μg/L 之间变化),在进行拟合时,这部分的误差不能忽略。而采用 TLS 方法在计算过程中同时考虑了参数误差和叶绿素 a 浓度的测量误差,得到了精度更高、更符合实际的拟合结果。

3.3 模型验证

随机选取 1/3 现场数据(81 组)对 LS 和 TLS 模型进行验证,由于叶绿素 a 的特征吸收峰为 440 nm 和 675 nm 附近,因此,选取 443 nm 和 675 nm 两个典型波长进行拟合,得出色素吸收系数的估算值,并与实测吸收系数进行了对比,结果如图 7 所示。图 7(a)为 443 nm 拟合结果,图 7(b)为 675 nm 拟合结果,图中红色实线为 TLS 拟合效果,绿色虚线为 LS 拟合效果。从图中可以看出,TLS 拟合效果优于 LS 拟合效果,原因在于 TLS 和 LS 的计算原理不同,TLS 同时考虑了自变量和因变量(吸收系数与叶绿素 a 浓度)的随机误差,减小了对估算结果的影响;675 nm 拟合效果优于 443 nm 拟合效果,分析原因为 675 nm 波段的吸收几乎完全都来自叶绿素 a 的贡献,其他色素在此波段吸收较小,而在 440 nm 波段,除叶绿素 a 有吸收,大部分的吸收都来自其他辅助色素。



(a) 443 nm 拟合结果 (b) 675 nm 拟合结果

图 7 典型波长实测 $a_{ph}(\lambda)$与 $a_{ph}^{M}(\lambda)$结果

3.4 模型评估

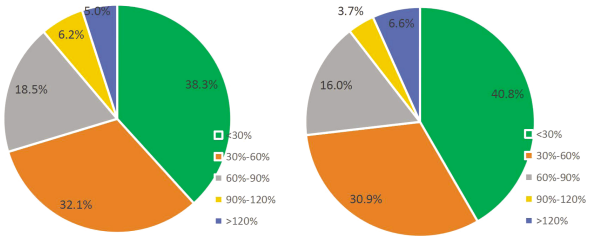
为了评价浮游植物色素吸收系数实测值 $a_{ph}(\lambda_i)$与模型估算值 $a_{ph}^{M}(\lambda_i)$之间的误差,分别采用平均绝对百分误差(Mean Absolute Percentage Error,*MAPE*)和均方根误差(Root Mean Square Error,*RMSE*)来评估模型估算效果。

3.4.1 *MAPE* 评估效果 如式(4)所示,光谱范围

400~700 nm,模型估算吸收系数值 $a_{ph}^{M}(\lambda)$与吸收系数实测值 $a_{ph}(\lambda_i)$之间的平均绝对百分误差表示为:

$$M=avg_{i=400}^{700}\left(\frac{\left|a_{ph}(\lambda_i)-a_{ph}^{M}(\lambda_i)\right|}{a_{ph}(\lambda_i)}\right) \tag{4}$$

根据式(4),计算评估参数 M,图 8 为 81 个模型验证站位的 M 分布图,图 8(a)和图 8(b)分别为 LS 和 TLS 结果。采用 LS 拟合得出的 $a_{ph}^{M}(\lambda)$,400~700 nm 光谱之间的平均偏差有 38.3%的站位数据在 30%以内,70.4%的站位数据 M 值在 60%之内;采用 TLS 拟合得出的 $a_{ph}^{M}(\lambda)$,400~700 nm 光谱之间的平均偏差有 40.8%的站位数据在 30%以内,71.7%的站位数据在 60%之内。由此可知,采用 TLS 获得的结果优于采用 LS 得到的结果。

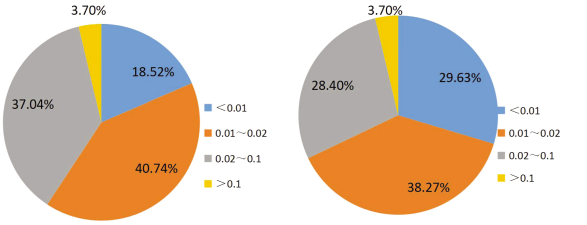


(a) LS 结果　　(b) TLS 结果

图 8　评估参数 M 分布图

3.4.2　*RMSE* 评估效果　均方根误差(*RMSE*)反映的是实测值与真值之间的偏离程度,R 值越小,二者之间的偏离程度越小,计算公式如式(5)所示,计算结果如图 9,图 9(a)和图 9(b)分别为 LS 和 TLS 计算结果。

$$R=\sqrt{avg_{i=400}^{700}\left(a_{ph}(\lambda_i)-a_{ph}^{M}(\lambda_i)\right)^2} \tag{5}$$



(a) LS 结果　　(b) TLS 结果

图 9　评估参数 R 分布图

从图中可知,400~700 nm 光谱范围内,LS 拟合得出的色素吸收系数 $a_{ph}^{M}(\lambda)$与实测 $a_{ph}(\lambda_i)$的均方根误差有 18.5%的数据小于 0.01,有 59.3%的数据控制在 0.02 内;采用 TLS 拟合得出的结果,有 29.6%的数据小于 0.01,67.9%的数据在 0.02 以内,说明 TLS 拟合效果优于 TL 拟合效果。

MAPE 为平均绝对百分比误差,范围[0,∞),*MAPE* 为 0,表示模型完美,*MAPE* 大于 100%,则表示劣质模型;*RMSE* 为均方根误差,范围[0,∞),*RMSE* 数值越大,说明实测值与真值之间的差值越大;两种评估方法均可反映出测量值与估算值之间的误差,两种方法都得出同样的结果,即 TLS 拟合结果优于 LS 拟合结果,但从结果来看,*RMSE* 的评估效果更好。

4　结 论

本文利用黄东海海域 2003-2012 年 6 个航次的 243 组现场实测数据,分别采用常规最小二乘法和总体最小二乘法进行了典型波段拟合,建立了浮游植物色素吸收系数估算模型,并加以验证和效果评估,得出以下结论。

(1) 采用 LS 和 TLS 方法计算典型波长的参数 $A(\lambda)$和指数 $E(\lambda)$。对于参数 $A(\lambda)$,两种方法在 400~700 nm 波段范围内变化趋势一致,TLS 结果稍大于 LS 结果;对于指数 $E(\lambda)$,400~700 nm 波段范围内变化趋势基本一致,但 TLS 数值明显大于 LS。与国外学者的研究结果比较,参数 $A(\lambda)$和指数 $E(\lambda)$都有相同的变化趋势,但数值略有不同,$A(\lambda)$偏高,$E(\lambda)$偏低。

(2) 利用 LS 和 TLS 模型估算 443 nm 和 675 nm 两个典型吸收波长的吸收系数,并与吸收系数实测值进行比较,结果显示,TLS 拟合结果优于 LS,675 nm 效果优于 443 nm。

(3) 定义模型评估参数 M 和 R,评价 LS 和 TLS 两种模型的拟合效果。结果显示,两种评估方法均能很好地说明,在 400~700 nm 光谱范围内,TLS 拟合结果优于 LS 拟合结果。

总之,总体最小二乘法在综合考虑了参数误差和测量误差的情况下,拟合结果精度更高,更符合实际情况,估算结果的准确性要优于常规最小二乘法,可将总体最小二乘法应用于浮游植物色素吸收系数的模型估算中。

参考文献：

[1] 朱建华，李铜基.黄东海海区浮游植物色素吸收系数与叶绿素 a 浓度关系研究[J].海洋技术，2004，23(4)：117-122.
[2] 王林，赵冬至，杨建洪，等.北黄海浮游植物吸收系数的季节性变化研究[J].海洋环境科学，2010，29(1)：60-65.
[3] Bricaud A，Babin M，Morel A，et al.Variability in the chlorophy Ⅱ -specific absorption coefficients of natural phytoplankton：Analysis and parameterization[J]. J Geophys Res，1995，100(C7)：13321-13332.
[4] Bricaud A，Morel A, Babin M，et al. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models[J]. J Geophys Res，1998，103：31033-31044.
[5] Matsuoka A，Huot Y，Shimada K，et al. Bio-optical characteristics of the western Arctic Ocean：implications for ocean color algorithms [J]. Canadian Journal of Remote Sensing，2007，33(6)：503-518.
[6] Churilova T，Suslin V, Krivenko O，et al. Light absorption by phytoplankton in the upper mixed layer of the Black Sea：Seasonality and Parametrization[J]. Frontiers in Marine Science，2017(4)：90.
[7] Meler J，Woźniak S B, Stoń-Egiert J，et al. Parameterization of phytoplankton spectral absorption coefficients in the Baltic Sea：General, monthly and two-component variants of approximation formulas[J]. Ocean Science，2018，14(6)：1523-1545.
[8] 曹文熙，杨跃忠，许晓强，等.珠江口悬浮颗粒物的吸收光谱及其区域模式[J]. 科学通报，2003，48(17)：1876-1882.
[9] 高庚.最小二乘法和总体最小二乘法线性回归中的估值漂移及其判定[D]. 太原：太原理工大学，2015.
[10] 鲁铁定.总体最小二乘平差理论及其在测绘数据处理中的应用[J]. 测绘学报，2013，42(4)：630.
[11] Mitchell B G，et al. Ocean optics protocols for satelite corlor sensor validation [M]. NASA/TM, 2002：231-257，269-283.

Application of Total Least Square Method in the Estimation of Pigment Absorption Coefficient Model of Phytoplankton

LI Zhi-feng, LI Tong-ji, ZHU Jian-hua
National Ocean Technology Center，*Tianjin* 300112，*China*

Abstract：The relationship between phytoplankton pigment absorption coefficient and the concentration of chlorophyll a was studied by applying the regular least square (LS) method and total least square (TLS) method. The LS and TLS models were established based on 243 sets of measured data obtained from 6 cruises during the period from 2003 to 2012. The pigment absorption coefficient values of typical bands were estimated by using the models, and the two models were verified. Two methods, mean absolute percentage error and root mean square error, were used to evaluate the model effect. The results showed that both evaluation methods can well explain that TLS fitting results are better than LS fitting results in the spectrum range of 400-700 nm. When parameter error and measurement error are taken into account in a comprehensive way, the overall least square method is more accurate and more consistent with the actual situation, with the accuracy of estimated results being better than those obtained through conventional least square method.

Key words：pigment absorption coefficient；chlorophyll a concentration；least square method；model；evaluation