顾及测量不确定性的水体悬浮物浓度遥感定量反演方法

艾烨霜 沈永林

中国地质大学(武汉)信息工程学院, 湖北 武汉 430074

摘要 在遥感定量反演的地面同步实测环节中,人为因素、环境变化、条件限制等测量不确定性因素会不可避免地引入数据噪声,致使水体悬浮物浓度反演精度降低。为此,提出一种顾及测量不确定性的水体悬浮物浓度遥感定量反演方法,即自适应抽样一致性极限学习机(ASAC-ELM)算法。该算法结合了极限学习机(ELM)、随机抽样一致性(RANSAC)和 N 邻近点抽样一致性(NAPSAC)方法的优势与特点,利用参数维度自适应地选取 RANSAC 或 NAPSAC 算法进行参数估计,避免了 ELM 算法易受非零均值正态分布数据噪声影响的缺陷。ASAC-ELM 算法通过选取局内点(非噪声点)数据建立模型,可去除噪声数据的干扰,提升模型的精度与适应性。通过模拟多组不同数量级且服从非零均值正态分布的随机数,将加性噪声引入训练数据中,实现不同噪声比条件下对 ASAC-ELM 算法的检验,并与 ELM 算法、传统反向传播(BP)神经网络算法进行了对比。结果表明,不同噪声比条件下,ASAC-ELM 算法的水质悬浮物浓度反演精度高于 ELM 算法和传统 BP 神经网络算法,且反演结果稳定性较高。

关键词 海洋光学; 遥感定量反演; 测量不确定性; 悬浮物浓度; 极限学习机; 随机抽样一致性; N 邻近点抽样一致性

中图分类号 TP79; X87 **文献标识码** A

doi: 10.3788/AOS201636.0701002

Measurement Uncertainty-Aware Quantitative Remote Sensing Inversion to Retrieve Suspended Matter Concentration in Inland Water

Ai Yeshuang Shen Yonglin

College of Information Engineering, China University of Geosciences, Wuhan, Hubei 430074, *China*

Abstract In the process of synchronous ground observation for quantitative remote sensing inversion, measurement uncertainty factors like human subjective factor, environmental change and condition restriction will induce data noise inevitably, which degrades the retrieval accuracy of the suspended matter concentration. Therefore, a measurement uncertainty-aware retrieval method named as the adaptive sample consensus extreme learning machine (ASAC-ELM) is proposed. ASAC-ELM integrates the merits of extreme learning machine (ELM), random sample consensus (RANSAC) and N adjacent points sample consensus (NAPSAC). The algorithm adaptively selects RANSAC or NAPSAC to estimate model parameters with the guidance of the parameter dimension, which avoids the problem that the ELM algorithm is sensitive to the non-zero normal distributed data noise. The ASAC-ELM algorithm selects inlying points (non-noise points) for model construction, thus can remove the interference from noise, and enhance the accuracy and flexibility of the model. In order to investigate the effectiveness of the proposed method under different noise conditions, a series of additive noise with non-zero mean normal distribution is introduced in the training data. The comparison among ASAC-ELM, ELM and traditional back propagation (BP) neural network algorithms is also conducted. The results show that for the retrieval of inland water suspended matter concentration under various noise conditions, the inversion accuracy and stability of ASAC-ELM is higher than those of ELM and the traditional BP neural network.

Key words oceanic optics; quantitative remote sensing inversion; measurement uncertainty; suspended matter

收稿日期: 2016-01-14; **收到修改稿日期**: 2016-02-29

基金项目: 国家自然科学基金(41501459,41301380)

作者简介: 艾烨霜(1993—),男,硕士研究生,主要从事水质遥感反演等方面的研究。E-mail: 13007163487@163.com

导师简介: 沈永林(1983—),男,博士,讲师,主要从事高光谱应用等方面的研究。E-mail: yonglinshen@gmail.com (通信联系人)

concentration; extreme learning machine; random sample consensus; N adjacent points sample consensus
OCIS codes 010.0280; 010.4450; 010.7340

1 引　言

悬浮物浓度是评价水体透明度和富营养化的重要参数，对水质和水环境监测及治理具有重要意义[1-3]。遥感技术具有快速、准确的优势，在水质监测中发挥重要作用[4-6]。美国陆地卫星(Landsat)系列广泛应用于水体总悬浮物(TSM)、无机悬浮物(ISM)等水质指标的预测。Olmanson 等[7]利用 Landsat 数据获取了明尼苏达州 10000 个湖体的水体透明度分布情况。Tebbs 等[8]利用 Landsat ETM+(增强型专题制图仪)数据反演叶绿素 a 浓度，反映了博戈里亚湖中高生物量的蓝藻水华。

水质遥感定量反演多采用线性回归模型。Olmanson 等[9]利用一元线性方程反演了密西西比河的水体质量评价指标。但线性模型在描述遥感观测指标与水质参量间的复杂关系方面稍显不足，易导致信息丢失和失真，实用性受限。因此人工神经网络(ANN)、极限学习机(ELM)等非线性模型逐渐被采用。孙德勇等[10]利用 ANN 对太湖水体悬浮物浓度进行了遥感反演尝试。尽管 ANN 能逼近任意复杂的非线性关系，但存在学习收敛速度慢、易陷入局部极值、网络结构难以确定等问题。ELM 作为一种新型单隐层前向神经网络，兼具 ANN 描述非线性关系的优点及自身网络结构易确定、学习速度快、泛化能力强等优势[11]。但在遥感定量反演的地面同步实测环节中，受测量方法、天气条件、实验员对操作规范的熟练程度等测量不确定性因素的影响，ELM 算法的实用性受到一定限制。

测量不确定性主要表现为 4 个方面：1) 悬浮物浓度在垂直方向的分布存在层化效应，采集的样品不能准确反映采样点处悬浮物浓度的实际情况[12]；2) 由于水质采样是接触性的，当船舶接近采样水域时，船舶扰动势必导致水面波动且带动湖水底部物质上浮，严重影响水体悬浮物浓度采样的准确性[12-13]；3) 悬浮物浓度的测量需在实验室利用化学化验等方式完成，在该过程中，悬浮物样品的降解及实验室内精密度、准确度的不同控制程度均会带来测量误差[14]；4) 在采样及化验过程中，仪器装置的系统误差、偶然因素等随机误差，实验人员的主观因素，在时空上具有极大不稳定性的客观自然因素(例如风)导致的粗差，也会使悬浮物浓度的测量存在较大不确定性[15]。综上所述，悬浮物浓度获取过程中的诸多测量不确定性均会导致较大的测量误差，而这些测量误差会影响数据分析和建模结果，进而导致遥感反演的悬浮物浓度结果不能客观地反映水体悬浮物浓度的空间分布状况。

针对数据测量不确定性影响水质反演精度的问题，众多学者开展了大量探索。Rousseeuw 等[16]利用数值计算方法研究并探讨了数据测量误差对最小二乘法求解结果的影响。研究表明，当误差服从零均值的正态分布时，最小二乘法可获得较理想的拟合结果；但若测量误差服从非零均值的正态分布或非正态分布时，该方法不可靠。由于水体悬浮物浓度地面同步测量中的不确定性因素众多且复杂，使得测量误差更倾向于服从非零均值的正态分布。因此，线性回归、传统反向传播(BP)神经网络以及利用最小二乘法进行参数估计的 ELM 算法均无法适用于该噪声条件，会导致悬浮物浓度反演精度低，无法满足业务应用要求。

本文提出一种顾及测量不确定性的水体悬浮物浓度遥感定量反演方法，即自适应抽样一致性极限学习机(ASAC-ELM)算法。该算法强化了模型对测量不确定性的适用性，能排除采集样本中数据噪声对反演精度的影响。模拟研究以太湖为研究区，以服从非零均值正态分布的随机数作为测量误差，以加性策略引入训练数据中，获得不同噪声比的悬浮物浓度数据，实现对算法的验证，并测试算法对噪声条件的适应性，以期辅助水质和水环境监测及治理等应用。

2 研究区及数据

研究区为长江中下游五大淡水湖之一的太湖，位于江苏省南部。太湖周边的环湖河流众多，入湖河流 23 条，出湖河流 9 条。其中，江苏段入湖河流 16 条，主要流入太湖梅梁湾、竺山湾、贡湖湾和西部沿岸区；浙江段入湖河流 7 条，主要流入太湖南部沿岸区。近年来，随着太湖流域周边地区经济的快速发展，流域内的土地利用类型不断发生变化，大量陆上泥沙和营养盐随着地表径流和降雨被带入湖中，导致湖体悬浮物的分布呈现空间上的差异。

研究以 Landsat 8 OLI(陆地成像仪)影像为数据源。该成像仪包含 9 个波段,8 个空间分辨率为 30 m 的多光谱段和 1 个分辨率为 15 m 的全色波段。实验共进行 2 次空地同步观测,各获取 1 景遥感影像,时间分别为 2013 年 8 月 4 日和 2014 年 8 月 7 日。经大气校正、几何校正、投影变换、区域裁剪等预处理,最终获得遥感影像的反射率数据。同时,在太湖水域采集了 80 个样本(图 1),其中 49 个采样点的数据获取时间为 2014 年 8 月 7 日(图 1 红色圆形标示点),另外 31 个采样点的数据获取时间为 2013 年 8 月 4 日(图 1 绿色方形标示点)。对采集的悬浮物数据进行统一量纲处理,获得采样点水体的 TSM 和 ISM 数据。其中,TSM 的均值为 44.97 mg/L,取值范围为[5.4,143.88];ISM 的均值为 31.17 mg/L,取值范围为[1.87,121.31]。

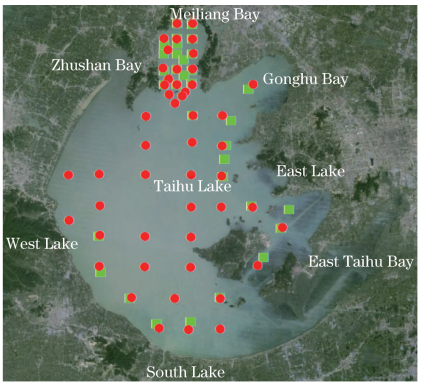

图 1　研究区及采样点的空间分布

Fig. 1　Spatial distribution of research area and sampling locations

3　自适应抽样一致性极限学习机

ASAC-ELM 算法结合了 ELM 算法、随机抽样一致性(RANSAC)算法[17]和 N 邻近点抽样一致性(NAPSAC)算法[18]的优势及特点。算法主要由 2 部分构成:1)变量间复杂的非线性关系到线性的映射;2)模型的自适应选择及参数估计。

3.1　变量间复杂的非线性关系到线性的映射

以单隐层网络结构为基础,假设输入层的输入变量个数为 n,隐含层神经元个数为 l,输出层的输出变量个数为 m,则通过随机给定的小数量级值来确定输入层与隐含层间的连接权值 $\boldsymbol{\omega}$ 和隐含层神经元阈值 $\boldsymbol{b}$,且 $\boldsymbol{\omega}$ 和 $\boldsymbol{b}$ 均在模型训练过程中保持不变。隐含层与输出层间的连接权值 $\boldsymbol{\beta}$ 为需要确定的参数[19-23],其中 $\boldsymbol{\omega}$、$\boldsymbol{\beta}$、$\boldsymbol{b}$ 的描述如下:

$$\boldsymbol{\omega}=\begin{bmatrix}\omega_{11} & \cdots & \omega_{1n}\\ \omega_{21} & \cdots & \omega_{2n}\\ \vdots & \ddots & \vdots\\ \omega_{l1} & \cdots & \omega_{ln}\end{bmatrix}_{l\times n},\ \boldsymbol{b}=\begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_l\end{bmatrix}_{l\times 1},\ \boldsymbol{\beta}=\begin{bmatrix}\beta_{11} & \cdots & \beta_{1m}\\ \beta_{21} & \cdots & \beta_{2m}\\ \vdots & \ddots & \vdots\\ \beta_{l1} & \cdots & \beta_{lm}\end{bmatrix}_{l\times m},\tag{1}$$

式中 ω_{rs} 为输入层第 r 个变量与隐含层第 s 个神经元间的连接权值,其中 $\boldsymbol{\omega}$ 第 i 行对应的向量 $\boldsymbol{\omega}_i=(\omega_{i1},\omega_{i2},\cdots,\omega_{in})^{\mathrm{T}}$;$b_i$ 为隐含层第 i 个神经元的阈值;β_{st} 为隐含层第 s 个神经元与输出层第 t 个变量间的连接权值,其中 $\boldsymbol{\beta}$ 矩阵第 j 行对应的向量 $\boldsymbol{\beta}_j=(\beta_{j1},\beta_{j2},\cdots,\beta_{jm})^{\mathrm{T}}$。

当样本数据集中样本个数为 N,输入值 $\boldsymbol{x}_j=(x_{j1},x_{j2},\cdots,x_{jn})^{\mathrm{T}}$,输出值 $\boldsymbol{t}_j=(t_{j1},t_{j2},\cdots,t_{jm})^{\mathrm{T}}$,给定隐含层神经元的激励函数 $g(x)$ 时,遥感观测指标与地面实测参量间的复杂关系可被描述为存在 $\boldsymbol{\omega}$、$\boldsymbol{\beta}$、$\boldsymbol{b}$ 使得

$$\sum_{i=1}^{l} g(\boldsymbol{\omega}_i\cdot\boldsymbol{x}_j+b_i)\boldsymbol{\beta}_i=\boldsymbol{t}_j,\ j=1,2,\cdots,N。\tag{2}$$

采用矩阵形式,(2)式可表示为

$$\boldsymbol{H\beta}=\boldsymbol{T},\tag{3}$$

式中 $\boldsymbol{H}$ 为隐含层输出矩阵,$\boldsymbol{T}$ 为输出层矩阵。

根据(1)~(3)式可实现变量间复杂的非线性关系到线性的映射,并得到相应的参数方程,为后续模型的确定和参数估计奠定基础。

3.2　模型的自适应选择及参数估计

ELM 算法采用最小二乘法作为参数的估计方法，但最小二乘法易受样本中非零均值正态分布数据噪声的影响。RANSAC 和 NAPSAC 算法能有效消除噪声干扰，但这两种算法在进行 ELM 参数估计时受模型复杂度和精度的制约。针对该问题，本文提出一种利用参数 $\boldsymbol{\beta}$ 的维度自适应选取模型参数估计算法的策略。即当模型参数 $\boldsymbol{\beta}$ 为低维时，选择 RANSAC 算法作为模型参数的估计算法；当模型参数 $\boldsymbol{\beta}$ 为高维时，选择 NAPSAC 算法。这主要是因为参数维度较低时，RANSAC 和 NAPSAC 算法达到相同精度所需的循环次数相近，但 RANSAC 算法的复杂度更低；而当参数维度较高时，NAPSAC 算法针对高维参数求解采取了邻近点具有特征一致性的原理，降低了参数初始化过程中噪声点引入的概率，与 RANSAC 算法相比其迭代次数大大减少，具有更高的模型学习效率。故通过自适应选择模型参数估计算法（RANSAC 算法或 NAPSAC 算法）对模型参数 $\boldsymbol{\beta}$ 进行求解，使得模型不仅能很好地适应噪声条件，且能顾及模型的学习效率，强化了模型的适应能力。参数 $\boldsymbol{\beta}$ 的维度判定和估计的具体步骤为：

1）利用 RANSAC 算法对（3）式中的参数 $\boldsymbol{\beta}$ 进行初始化估计，即设定算法最大的循环学习次数 n（如 $n=1000$），并记录 RANSAC 算法正确估计模型参数时所需的迭代次数，记为 c；

2）对 c 的大小进行判定，若 $c<n$，则模型参数 $\boldsymbol{\beta}$ 的维度属于低维，选择 RANSAC 算法估计的参数结果有效，终止学习进程；否则，模型参数 $\boldsymbol{\beta}$ 的维度属于高维，RANSAC 算法失效，并利用 NAPSAC 算法对模型参数重新进行估计，直至模型参数被正确估计，循环结束。

ASAC-ELM 算法结合了 ELM 算法、RANSAC 算法和 NAPSAC 算法各自的优势特点，能有效去除数据中的噪声（服从非零均值或零均值的正态分布）影响，增强算法对于测量不确定性因素的适应性，并有望提升算法在水体悬浮物浓度遥感定量反演应用中的适用性和泛化能力。

4　实验设计

4.1　模型及数据设定

悬浮物浓度的变化有其特有的光谱反射率特性，这为基于 Landsat 8 数据的水体悬浮物浓度反演的波段选择提供了理论依据。Sang 等[24]基于米氏散射机理分析了不同尺寸悬浮颗粒物在 200～1000 nm 波段内的散射特性；陈亚慧等[25]发现蓝绿波段（470～575 nm）能较好地反映悬浮物的粒径大小；刘忠华等[26]在模拟太湖颗粒物后向散射特征时发现 560 nm 处颗粒物后向散射系数与悬浮物浓度相关性较高。Shi 等[27]利用中分辨率成像光谱仪（MODIS）数据反演太湖水体悬浮物浓度时发现 645 nm 处光谱强度与悬浮物浓度具有较高的相关性。Wang 等[28-29]利用 MODIS 数据进行太湖水质指标检测与评估时发现近红外波段（748～869 nm）和短波红外波段（1240～2130 nm）能较好地反映浑浊度较高的太湖水体的内部光学和生物光学特征。考虑到本文所采用的遥感定量反演模型本质上是一种网络结构模型，对输入变量存在加权运算，故以 Landsat 8 的 1～7 波段（433～2300 nm）作为模型输入。该波段选择策略一方面能保证信息的充分利用，另一方面能降低冗余信息的干扰[30-31]。对获取的最终样本数据，均匀抽取 15％的样本作为测试数据（记作 II），剩余的 85％作为训练数据（记作 I）。为验证不同噪声条件下的稳健性，考虑到水质测量不确定性产生的原因，设定测量误差服从非零均值的正态分布；通过模拟服从非零均值正态分布的随机数，以加性策略引入训练样本中，以获得不同噪声比的训练数据。模拟研究中，利用正态分布随机数产生函数（$\mu=100$，$\sigma^2=30$）随机产生 4 类不同噪声比的随机数，每类随机数为 100 组，相应地获得 4 类不同噪声比的训练数据，分别记为 I_{1i}、I_{2i}、I_{3i}、I_{4i}，其中 $i=1,2,\cdots,100$，如表 1 所示。其中，最大的噪声比限制在 26.58％以内，保证了测量误差服从正态分布时模型的反演精度在±30％之内，更符合实际应用需求[12]。

模拟中分别用带有不同噪声比的训练数据（I_{1i}、I_{2i}、I_{3i}、I_{4i}）进行模型训练，然后利用测试数据 II 进行测试。将每一类训练数据对应的计算结果取平均作为最终结果，以此来消除随机性带来的偶然误差。

4.2　模型参数设置

在 ASAC-ELM 算法中，神经元个数、激励函数、迭代终止条件、局内点判定阈值均为影响模型精度的重要参数。本文 ASAC-ELM 算法以反曲函数（sigmoidal）作为激励函数，其迭代终止条件和局内点判定阈值的设定与训练数据噪声比密切相关，故其值设定主要参考该指标（表 1）。

表 1　模拟的噪声数据和模型参数

Table 1　Description of simulated noise data and model parameters

Data	Noise parameter		Model parameter	
	Number of noise points	Ratio of noise /%	Iteration stopping criteria	Inlying point threshold
I_{1i}	0	0	60	30
I_{2i}	5	7	55	30
I_{3i}	10	15	50	30
I_{4i}	15	22	45	30

神经元个数的选择相对复杂。模拟中采取的策略是对给定的测试数据依次增加模型的神经元个数，然后根据测试精度（采用标准误差来衡量）的最优值选择合适的神经元个数。具体步骤为：针对未加噪声的训练数据，将神经元个数由 1 到 40、以 1 为间隔依次递增，分别计算其对应的测试精度；当测试精度最优时，选择与之对应的神经元个数作为模型最终的神经元个数。用未加噪声的训练数据确定神经元个数，并将确定的最优神经元个数作为其余带有不同噪声比训练数据进行模型训练时的参数，目的是使训练数据中的噪声比成为整个模拟过程中唯一的变量。通过太湖数据计算可知，当神经元个数达到一定值时，继续增加并不能提升测试精度，反而会降低精度，如图 2 所示。当神经元个数处于[1，3]区间时，随着个数的增加 TSM 和 ISM 两种悬浮物浓度指标的测试精度也随之提高；当神经元个数处于[3，15]区间时，测试精度较低，达到相对稳定的水平；当神经元个数处于 [15，40] 区间时，随着个数的增加，TSM 和 ISM 的测试精度均降低。为此，将 ASAC-ELM 模型中 TSM 和 ISM 指标测试的最优神经元个数分别设置为 3 个和 4 个。

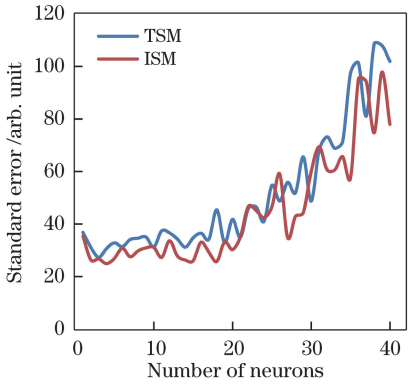

图 2　不同神经元个数下的模型测试精度

Fig. 2　Statistical summary of testing accuracy with different number of neurons

另外，将 ASAC-ELM 算法与 ELM 算法、传统 BP 神经网络模型进行对比。为保证算法的可比性，ELM 算法神经元个数和激励函数等参数的确定与 ASAC-ELM 算法保持一致。对于传统 BP 神经网络模型，根据 Kolmogorv 定理[32]，采用逐步增长法确定最优神经元个数为 5，隐含层和输出层传递函数分别为 tansig 和 purelin，训练函数为 traingdx，期望误差为 0.0022，最大训练循环次数为 6000 次。

4.3　精度评估

利用散点图、残差分布曲线（残差取绝对值）以及标准误差（f_{RMSE}）定量评价模型的测试精度。f_{RMSE} 可表示为

$$f_{\mathrm{RMSE}}=\sqrt{\sum_{i=1}^{n}(x_i-x_i')\Big/n}\,, \tag{4}$$

式中 x_i 为预测值，x_i' 为测量值，n 为样本个数。测试精度越高，f_{RMSE} 值越小，散点图中的点越靠近直线 $y=x$，残差分布曲线越靠近直线 $y=0$。通过定量评价，对比不同噪声比数据下模型的噪声适应性能力，验证 ASAC-ELM 算法的有效性。

5　结果及讨论

5.1　精度对比及分析

利用 4 类带有不同噪声比的训练数据 I_1、I_2、I_3、I_4，对 ASAC-ELM 算法、ELM 算法以及传统的 BP 神

经网络进行训练，并用测试数据 *II* 进行测试，通过上文介绍的定量评估方法进行精度评价，分别获得针对 TSM 和 ISM 两种悬浮物浓度指标的各算法在不同噪声条件下的散点图（图 3、4）、残差分布曲线（图 5、6）和 f_{RMSE} 值（表 2）。

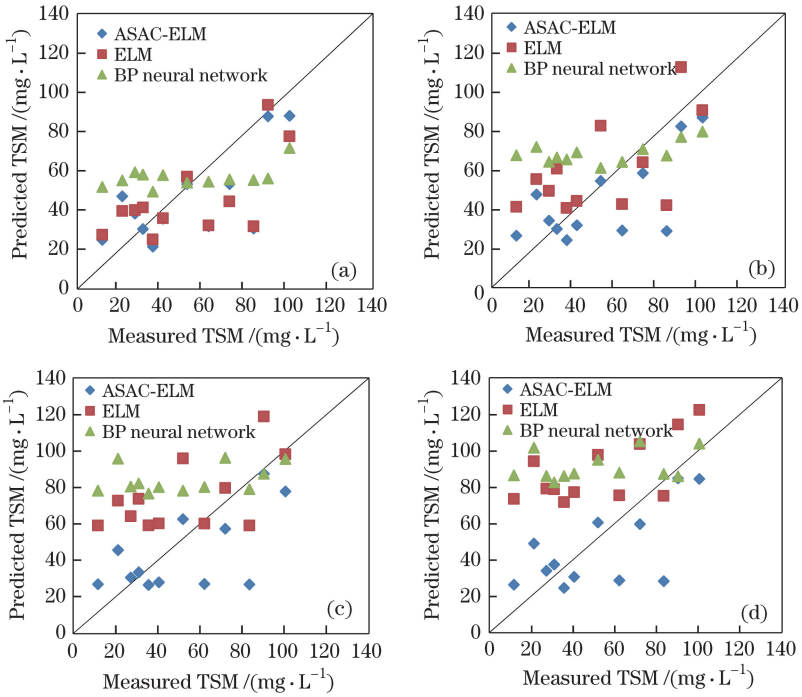

图 3　TSM 实测值和预测值的散点图。(a) 0 噪声点；(b) 5 噪声点；(c) 10 噪声点；(d) 15 噪声点

Fig. 3　Scatter-plots of measured and predicted TSM. (a) 0 noise point; (b) 5 noise points; (c) 10 noise points; (d) 15 noise points

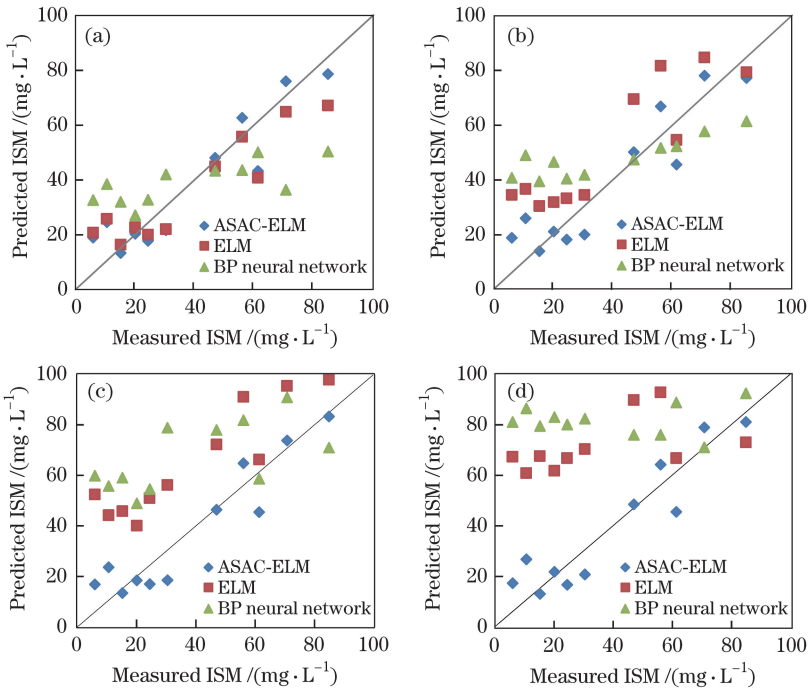

图 4　ISM 实测值和预测值的散点图。(a) 0 噪声点；(b) 5 噪声点；(c) 10 噪声点；(d) 15 噪声点

Fig. 4　Scatter-plots of measured and predicted ISM. (a) 0 noise point; (b) 5 noise points; (c) 10 noise points; (d) 15 noise points

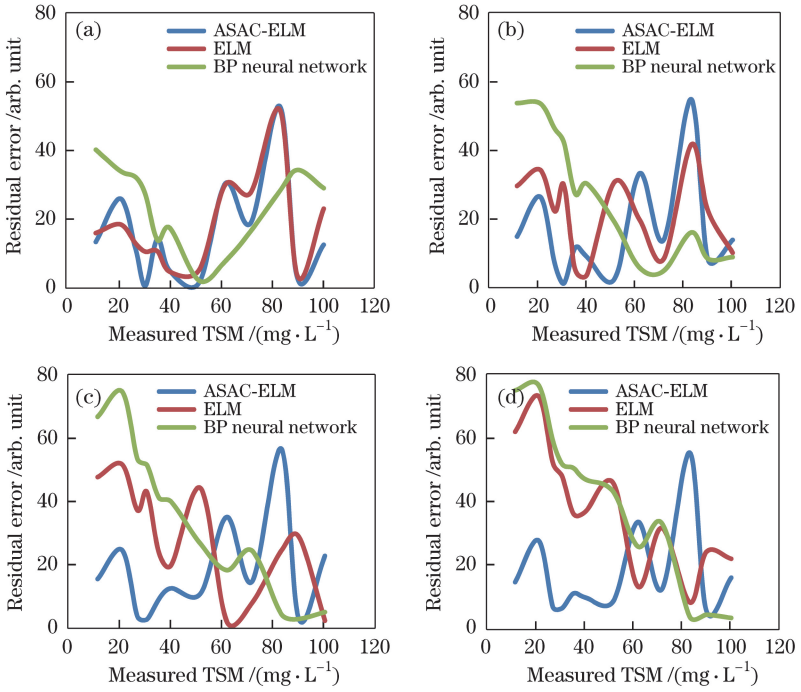

图 5　TSM 实测值和预测值的残差。(a) 0 噪声点;(b) 5 噪声点;(c) 10 噪声点;(d) 15 噪声点

Fig. 5　Residual error of measured and predicted TSM. (a) 0 noise point; (b) 5 noise points; (c) 10 noise points; (d) 15 noise points

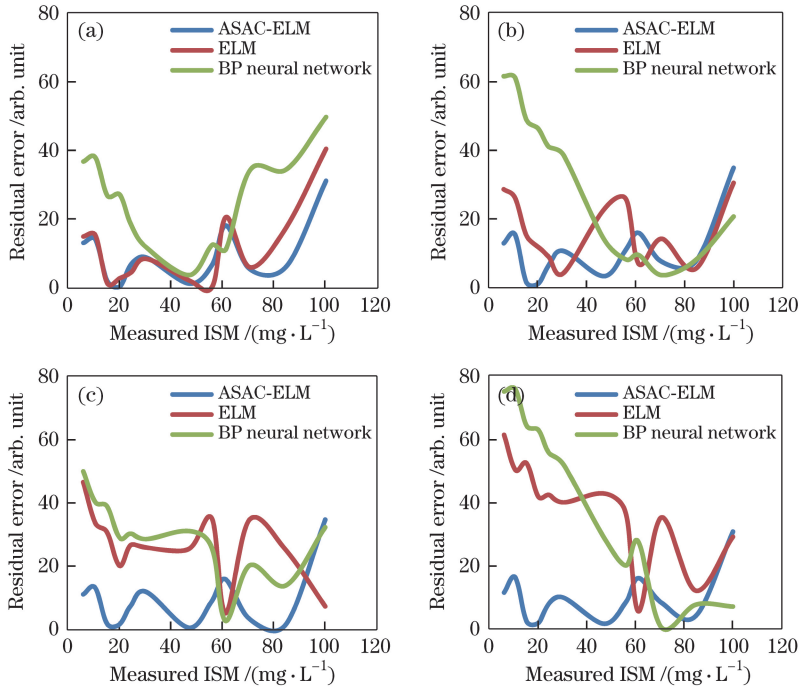

图 6　ISM 实测值和预测值的残差。(a) 0 噪声点;(b) 5 噪声点;(c) 10 噪声点;(d) 15 噪声点

Fig. 6　Residual error of measured and predicted ISM. (a) 0 noise point; (b) 5 noise points; (c) 10 noise points; (d) 15 noise points

由散点图 3(a)和图 4(a)、残差分布曲线图 5(a)和图 6(a)可知,当训练数据中未引入噪声数据时,ASAC-ELM 和 ELM 算法的散点在直线 $y=x$ 附近分布较为相似,残差曲线也几乎重叠,平均偏离量(取 TSM 和 ISM 两指标的偏离平均值,下同)分别为 12.55 和 14.46,相差较小。表明在数据质量较好时,ASAC-ELM 算法和 ELM 算法具有相似的反演精度;由传统 BP 神经网络得到的散点分布和残差曲线结果

表 2　不同噪声条件下的 f_{RMSE} 值

Table 2　f_{RMSE} under different noise conditions

Algorithm	TSM				ISM			
	I_1	I_2	I_3	I_4	I_1	I_2	I_3	I_4
ASAC-ELM	22.27	23.44	23.91	23.28	13.40	14.91	14.71	14.00
ELM	22.68	31.67	36.47	43.09	16.21	24.10	31.08	41.96
BP neural network	25.39	36.28	44.94	49.28	19.53	30.14	37.05	46.18

显示，平均偏离量为 21.95，略差于 ASAC-ELM 和 ELM 算法结果。表明 ELM 算法在实际应用中的泛化能力比传统的 BP 神经网络更强。

比较图 3～6 中各列可得如下结果。1）当训练数据中引入服从非零均值正态分布的数据噪声时，ELM 算法和传统 BP 神经网络得到的散点图严重偏离直线 $y=x$，残差分布曲线也明显偏离 $y=0$，平均偏离量分别为 18.99 和 28.18，明显高于未引入数据噪声时的平均偏移量 14.46 和 21.95，精度衰退率为 31.3% 和 28.4%。表明 ELM 算法和传统 BP 神经网络受数据噪声的影响较大，且 ELM 算法的参数估计策略易受服从非零均值正态分布数据噪声的影响。2）当训练数据中噪声数据的比例（或数量）不断增加时，ELM 算法和传统 BP 神经网络得到的散点图偏离直线 $y=x$，且残差曲线偏离直线 $y=0$ 的偏离量也会随之增加，平均偏离量分别为 14.46、18.99、27.01、37.58 和 21.95、28.18、31.30、39.75。表明 ELM 算法和传统 BP 神经网络随着数据噪声的增加，其模型性能降低，测试精度也会随之降低。3）ASAC-ELM 算法的散点图分布与 ELM 算法和传统 BP 神经网络相比更加贴近直线 $y=x$，其残差分布曲线也更靠近直线 $y=0$。引入数据噪声前后平均偏离量分别为 12.55 和 13.43，明显低于 ELM 算法的 14.46 和 18.99 以及传统 BP 神经网络的 21.95 和 38.18。当训练数据中引入服从非零均值正态分布数据噪声的比例（或数量）不断增加时，ASAC-ELM 算法得到的散点分布和残差分布曲线的平均偏移量分别为 12.55、13.43、13.39、13.56，并未出现明显的增加。表明 ASAC-ELM 算法的噪声适应性和算法稳定性比 ELM 算法和传统 BP 神经网络更好。

由图 5 和图 6 可知，随着训练数据中引入噪声数据的比例（或数量）不断增加，ELM 算法和传统 BP 神经网络得到的残差分布曲线在坐标轴靠左部分（即测试数据 II 中值偏小的部分）的残差值会随之急剧增加；而 ASAC-ELM 算法在此区域的残差值并未明显增加，且较为稳定地靠近直线 $y=0$，表明 ASAC-ELM 算法对低值区域具有更好的反演能力。

从表 2 的各项 f_{RMSE} 值可知，对于悬浮物浓度指标 TSM 和 ISM，当训练数据中引入服从非零均值正态分布数据噪声的比例（或数量）不断增加时，ELM 算法和传统 BP 神经网络得到的 f_{RMSE} 值会随之增加，特别是当训练数据中噪声比由 0 增加至 7% 时，f_{RMSE} 出现明显的增加，增量分别为 39.6%、42.9%（TSM）和 48.7%、54.3%（ISM）。进一步表明 ASAC-ELM 算法相比 ELM 算法和传统 BP 神经网络更好地考虑了测量不确定性因素，对噪声条件下的悬浮物浓度反演具有更好的适应性。

5.2　太湖悬浮物浓度反演结果

针对太湖区域，利用 80 个样本点的 TSM、ISM 实测值（不含模拟噪声）及相应的光谱值确定 ASAC-ELM 算法、ELM 算法及传统 BP 神经网络模型的模型参数，然后将模型应用于 2013 年 8 月 4 日整个太湖水域的 TSM、ISM 定量推演，并制成太湖区域悬浮物浓度的专题图，如图 7 和图 8 所示。

由图 7 和图 8 可知，整个太湖区域的 TSM 和 ISM 分布呈现明显的东西差异。北部的梅梁湾、竺山区、湖心区、西部沿岸区以及西南部水域均表现出较高的悬浮物浓度，而东太湖、胥口湾等东部沿岸区的悬浮物浓度相对低于西部。湖心区由于湖面开阔，底泥受风浪搅动的影响较大，湖底物质上浮增加了水体中悬浮物的浓度，故湖心区的悬浮物浓度较高。西南湖区一方面由于入湖径流经过浙西山区，携带了大量的悬浮颗粒物进入湖体，另一方面由于此处水域开阔，易受到风浪的影响。北部的梅梁湾、竺山区也受入湖河流的影响，悬浮物浓度相对较高。东太湖由于地处一个狭长的湾，受风浪影响较小，且此处的沉水植物对沉积物再悬浮的抑制作用较强，使悬浮物浓度相对较低。同时东部沿岸区多为出湖河流，引入的悬浮物相对较少，浓度也相对较低。

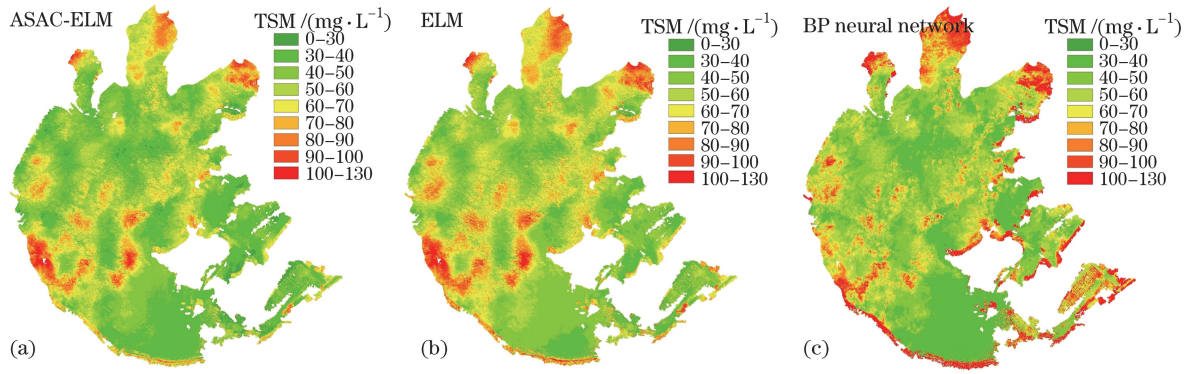

图 7　(a)ASAC-ELM 算法、(b)ELM 算法和(c)传统 BP 神经网络反演的 TSM 专题图

Fig. 7　TSM retrieval results by (a) ASAC-ELM algorithm, (b) ELM algorithm and (c) traditional BP neural network

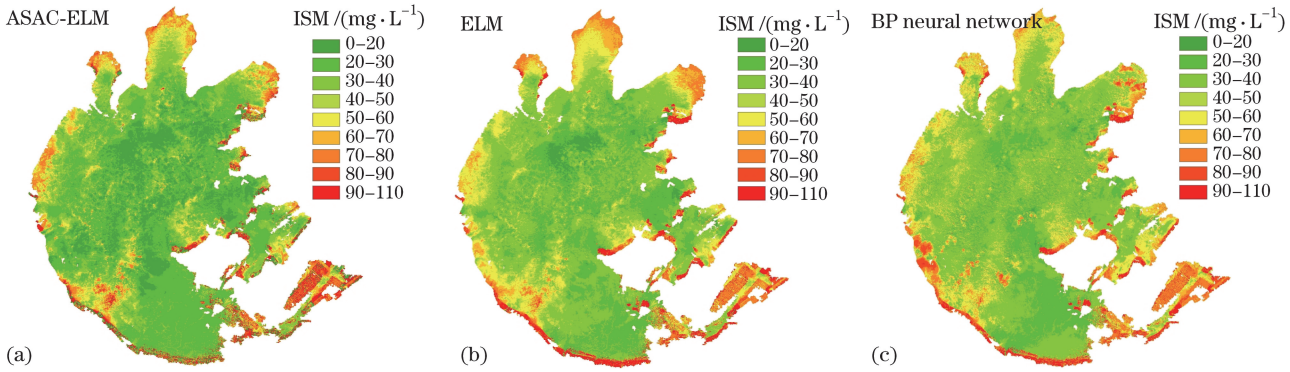

图 8　(a)ASAC-ELM 算法、(b)ELM 算法和(c)传统 BP 神经网络反演的 ISM 专题图

Fig. 8　ISM retrieval results by (a) ASAC-ELM algorithm, (b) ELM algorithm and (c) traditional BP neural network

根据 ASAC-ELM 算法、ELM 算法及传统 BP 神经网络得到的反演结果，三种算法得到的悬浮物浓度空间分布存在一定的差异。与 ASAC-ELM 算法和 ELM 算法相比，传统 BP 神经网络得到的悬浮物浓度空间分布从数值上较为跳跃，即从高浓度到低浓度的变化相对较为剧烈，特别是在边界区域表现得更加明显。表明传统 BP 神经网络在实际应用中受限于局部收敛等缺陷，其适用性比 ASAC-ELM 算法和 ELM 算法弱；同时，相对于 ASAC-ELM 算法，ELM 算法得到的悬浮物浓度在低浓度分布空间的具体值相对偏高，表明 ELM 算法精度易受测量不确定性引入的数据噪声的影响，ASAC-ELM 对悬浮物低浓度区的反演精度更高，对数据噪声的适应性更好。

6　结　　论

在水体悬浮物浓度反演的地面同步实测环节中，由测量不确定性导致的测量误差是不可避免的。常见的反演模型受数据噪声的影响，反演精度和适应性都面临较大的挑战。针对该问题，提出一种顾及测量不确定性的水体悬浮物遥感定量反演方法，即 ASAC-ELM 算法。该算法考虑了测量不确定性产生数据噪声的原因及 ELM 算法受数据噪声影响的环节。ASAC-ELM 算法能实现变量间复杂非线性关系到线性的映射，并有效探测和排除数据中的噪声点，选取局内点进行训练学习，强化算法的适应性并提升悬浮物浓度反演的精度。通过模拟不同噪声比训练数据对提出算法进行检验，发现 ASAC-ELM 算法对噪声的适应性明显优于 ELM 算法和传统 BP 神经网络。在太湖水域的水体悬浮物浓度反演中，ASAC-ELM 算法能适应噪声条件，自适应地选择参数估计算法，有效地排除了样本数据中噪声的影响，使得反演结果更为合理准确，从而满足水质遥感定量反演业务化应用需求。

模拟时假定数据误差主要来源于测量不确定性，但水体悬浮物浓度遥感定量反演中，除测量误差外，系统误差、大气影响及预处理等因素也会引入数据噪声，这是悬浮物浓度反演过程中引入噪声的另一重要原因，也是后续研究和探索的重要内容。

参 考 文 献

1 Shi Kun, Li Yunmei, Wang Qiao, *et al*.. Study of scattering coefficients model in inland eutrophic lake[J]. Acta Optica Sinica, 2010, 30(9): 2478-2485.
施 坤, 李云梅, 王 桥, 等. 内陆湖泊富营养化水体散射系数模型研究[J]. 光学学报, 2010, 30(9): 2478-2485.

2 Huang Changchun, Li Yunmei, Sun Deyong, *et al*.. Research of scattering spectrum characteristic and formative mechanism of Taihu lake waters[J]. Acta Optica Sinica, 2011, 31(5): 0501003.
黄昌春, 李云梅, 孙德勇, 等. 太湖水体散射光谱特性及其形成机理研究[J]. 光学学报, 2011, 31(5): 0501003.

3 Liu Yanfang, Su Rongguo, Zhou Qianqian, *et al*.. Rapid modeling offshore eutrophication technique using optical parameters of CDOM[J]. Chinese J Lasers, 2014, 41(12): 1215001.
刘艳芳, 苏荣国, 周倩倩, 等. 基于 CDOM 光学参数的近海富营养化快速评价技术[J]. 中国激光, 2014, 41(12): 1215001.

4 Verdin J P. Monitoring water quality conditions in large western reservoir with Landsat imagery[J]. Photogrammetric Engineering and Remote Sensing, 1985, 51(3): 343-353.

5 Wang M, Shi W, Tang J. Water property monitoring and assessment for China's inland Lake Taihu from MODIS-Aqua measurements[J]. Remote Sensing of Environment, 2011, 115(3): 841-854.

6 Randolph K, Wilson J, Tedesco L, *et al*.. Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin[J]. Remote Sensing of Environment, 2008, 112(11): 4009-4019.

7 Olmanson L G, Bauer M E, Brezonik P L. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes[J]. Remote Sensing of Environment, 2008, 112(11): 4086-4097.

8 Tebbs E J, Remedios J J, Harper D M. Remote sensing of chlorophyll-a as a measure of cyanobacterial biomass in Lake Bogoria, a hypertrophic, saline-alkaline, flamingo lake, using Landsat ETM+[J]. Remote Sensing of Environment, 2013, 135: 92-106.

9 Olmanson L G, Brezonik P L, Bauer M E. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The Mississippi River and its tributaries in Minnesota[J]. Remote Sensing of Environment, 2012, 130: 254-265.

10 Sun Deyong, Li Yunmei, Wang Qiao, *et al*.. Study on remote sensing estimation of suspended matter concentration based on *in situ* hyperspectral data in Lake Tai waters[J]. Journal of Infrared and Millimeter Wave, 2009, 28(2): 124-128.
孙德勇, 李云梅, 王 桥, 等. 基于实测高光谱的太湖水体悬浮物浓度遥感估算研究[J]. 红外与毫米波学报, 2009, 28(2): 124-128.

11 Cui Dongwen. Application of extreme learning machine to total phosphorus and total nitrogen forecast in lakes and reservoirs[J]. Water Resources Protection, 2013, 29(2): 61-66.
崔东文. 极限学习机在湖库总磷、总氮浓度预测中的应用[J]. 水资源保护, 2013, 29(2): 61-66.

12 Chen Jun, Fu Jun, Sun Jihong. The application of the numerical method to simulating the impact of the observation errors on the parameters of the water quality retrieval model: A case study of chlorophyll-a concentration[J]. Remote Sensing for Land & Resources, 2011, 23(1): 57-61.
陈 军, 付 军, 孙记红. 用数值方法模拟观测误差对水质浓度反演模型参数的影响——以叶绿素 a 浓度为例[J]. 国土资源遥感, 2011, 23(1): 57-61.

13 Giardino C, Bresciani M, Valentini E, *et al*.. Airborne hyperspectral data to assess suspended particulate matter and aquatic vegetation in a shallow and turbid lake[J]. Remote Sensing of Environment, 2015, 157: 48-57.

14 Wei Lihang, Hong Zhengfang. Estimation of measurement uncertainty in the field of environmental monitoring[J]. Environmental Science & Technology, 2014, 37(5): 176-181.
韦利杭, 洪正昉. 环境监测领域测量不确定度的评估[J]. 环境科学与技术, 2014, 37(5): 176-181.

15 Zhou Bin, Liu Wenqing, Qi Feng, *et al*.. Error analysis in differential optical absorption spectroscopy[J]. Acta Optica Sinica, 2002, 22(8): 957-961.
周 斌, 刘文清, 齐 锋, 等. 差分吸收光谱法测量大气污染的测量误差分析[J]. 光学学报, 2002, 22(8): 957-961.

16 Rousseeuw P J, Leroy A M. Robust regression and outlier detection[M]. San Francisco: John Wiley & Sons, 1987.

17 Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. Communication of the ACM, 1981, 24(6): 381-395.

18 Myatt D R, Torr P H S, Nasuto S J, *et al*.. NAPSAC: High noise, high dimensional robust estimation - it's in the bag [C]. Proceedings of British Machine Vision Conference, Cardiff, 2002, 2: 458-467.

19 Huang G B, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. IEEE Transactions on Neural Networks, 2006, 17(4): 879-892.

20 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: A new learning scheme of feed forward neural networks [C]. Proceedings of IEEE International Joint Conference on Neural Networks, Budapest, 2004, 2: 985-990.

21 Huang G B, Zhou H, Ding X, *et al*.. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2012, 42(2): 513-529.

22 Bartlett P L. For valid generalization, the size of the weights is more important than the size of the network[M]. // Advances in neural information processing systems. Cambridge: MIT Press, 1997: 134-140.

23 Bartlett P L. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of network[J]. IEEE Transactions on Information Theory, 1998, 44(2): 525-536.

24 Vo Quang Sang, Feng Peng, Tang Bin, *et al*.. Study on properties of light scattering based on Mie scattering theory for suspended particles in water[J]. Laser & Optoelectronics Progress, 2015, 52(1): 013001.
Vo Quang Sang, 冯　鹏, 汤　斌, 等. 基于米氏散射理论的水中悬浮颗粒物散射特性计算[J]. 激光与光电子学进展, 2015, 52(1): 013001.

25 Chen Yahui, Qiu Zhongfeng, Sun Deyong, *et al*.. Remote sensing of suspended particle size in Yellow Sea and Bohai Sea [J]. Acta Optica Sinica, 2015, 35(9): 0901008.
陈亚慧, 丘仲峰, 孙德勇, 等. 黄渤海悬浮颗粒物粒径的遥感反演研究[J]. 光学学报, 2015, 35(9): 0901008.

26 Liu Zhonghua, Li Yunmei, Tan Jing, *et al*.. Simulation of backscattering properties of particles in Taihu Lake based on optical closure principle[J]. Acta Optica Sinica, 2012, 32(7): 0701002.
刘忠华, 李云梅, 檀　静, 等. 基于光学闭合原理的太湖水体颗粒物后向散射特性模拟[J]. 光学学报, 2012, 32(7): 0701002.

27 Shi K, Zhang Y L, Zhu G W, *et al*.. Long-term remote monitoring of total suspended matter concentration in Lake Taihu using 250 m MODIS-Aqua data[J]. Remote Sensing of Environment, 2015, 164: 43-56.

28 Shi W, Wang M. Satellite observations of the seasonal sediment plume in central East China Sea[J]. Journal of Marine Systems, 2010, 82(4): 280-285.

29 Wang M. Remote sensing of the ocean contributions from ultraviolet to near-infrared using the shortwave infrared bands: Simulations[J]. Applied Optics, 2007, 46(9): 1535-1547.

30 Zhang Yuchao, Qian Xin, Qian Yu, *et al*.. Quantitative retrieval of chlorophyll a concentration in Taihu Lake using machine learning methods[J]. Environment Science, 2009, 30(5): 1321-1328.
张玉超, 钱　新, 钱　瑜, 等. 基于机器学习方法的太湖叶绿素 a 定量遥感研究[J]. 环境科学, 2009, 30(5): 1321-1328.

31 Wang Xiangyu, Wang Xili. Gray expansion combined GA-BP neural network model for water quality retrieving of Weihe River by remote sensing[J]. Remote Sensing Technology and Application, 2010, 25(2): 251-256.
王翔宇, 汪西莉. 结合灰色扩充的 GA-BP 神经网络模型在渭河水质遥感反演中的应用[J]. 遥感技术与应用, 2010, 25(2): 251-256.

32 Xue Pengsong, Feng Minquan, Xing Xiaopeng. Water quality prediction model based on Markov chain improving gray neural network[J]. Journal of Wuhan University (Engineering Edition), 2012, 45(3): 319-324.
薛鹏松, 冯民权, 邢肖鹏. 基于马尔科夫链改进灰色神经网络的水质预测模型[J]. 武汉大学学报(工学版), 2012, 45(3): 319-324.