

DOI: 10.7524/j.issn.0254-6108.2017.03.2016070104

王小平, 张飞, 于海洋, 等. 基于多元线性模型、支持向量机(SVM)模型和地统计方法的地表水溶解性总固体(TDS)估算及其精度对比——以艾比湖流域为例[J]. 环境化学, 2017, 36(3): 666-676.

WANG Xiaoping, ZHANG Fei, YU Haiyang, et al. Comparison of prediction accuracies of tds in the surface water in Ebinur Lake based on multivariate linear model, SVM model, and geostatistics method[J]. Environmental Chemistry, 2017, 36(3): 666-676.

## 基于多元线性模型、支持向量机(SVM)模型和地统计方法的地表水溶解性总固体(TDS)估算及其精度对比——以艾比湖流域为例\*

王小平<sup>1,2</sup> 张 飞<sup>1,2,3\*\*</sup> 于海洋<sup>1,2</sup> KUNG Hsiang-te<sup>4</sup>

(1. 新疆大学资源与环境科学学院, 乌鲁木齐, 830046; 2. 新疆大学绿洲生态教育部重点实验室, 乌鲁木齐, 830046; 3. 新疆智慧城市与环境建模普通高校重点实验室, 乌鲁木齐, 830046; 4. 美国孟菲斯大学地球科学系, 田纳西州 孟菲斯, 38152, 美国)

**摘 要** 地表水溶解性总固体(TDS)是地表水各组分浓度的总指标,是地表水水化学特性长期演变的最终结果,也是表征水文地球化学作用过程的重要参数,TDS的高低直接影响地表水的含盐量.本研究以艾比湖流域为研究对象,结合实测地表水 TDS 数据;选用准同步的 Landsat OLI 数据,首先,利用光谱诊断指数选取与地表水 TDS 相关性较高的波段,其次,利用地统计方法、多元线性回归模型和支持向量机(SVM)模型对 TDS 进行预测,并对其结果进行精度比较.结果表明,SVM 模型为最优估测模型,拟合决定系数  $R^2$  为 0.97,均方误差(RMSE)为 50.59;多元线性回归模型的精度与 SVM 模型精度较为接近,拟合决定系数  $R^2$  为 0.9, RMSE 为 66.55;地统计克里格插值法预测精度最低,拟合决定系数  $R^2$  为 0.87, RMSE 为 95.73.遥感估测 SVM 模型预测值在大区域能较好地反映出艾比湖流域 TDS 的总体特征.该模型在水质遥感领域的应用中具有良好的可行性和有效性,其预测结果也与艾比湖流域水体 TDS 的实际分布相吻合,因此遥感估测 SVM 模型在水质估测中具有一定的应用潜力.

**关键词** TDS, SVM 模型, 多元线性回归模型, 地统计, 艾比湖流域.

## Comparison of prediction accuracies of tds in the surface water in Ebinur Lake based on multivariate linear model, SVM model, and geostatistics method

WANG Xiaoping<sup>1,2</sup> ZHANG Fei<sup>1,2,3\*\*</sup> YU Haiyang<sup>1,2</sup> KUNG Hsiang-te<sup>4</sup>

(1. College of Resources and Environment Science, Xinjiang University, Urumqi, 830046, China;

2. Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi, 830046, China;

3. Key Laboratory of Xinjiang Wisdom City and Environment Modeling, Urumqi, 830046, China;

4. Department of Earth Sciences, University of Memphis, Memphis, TN 38152, USA)

**Abstract:** Total dissolved solids (TDS) of surface water is the total index of concentration of chemical components in surface water, and is an important parameter for the characterization of hydro-geochemical action. TDS directly affects the salinity of surface water. In this study, we chose Ebinur lake watershed as the study area, measured TDS data of the surface water, and analyzed

2016年7月1日收稿(Received: July 1, 2016).

\* 国家自然科学基金(41361045, 41130531)资助.

Supported by National Natural Science Foundation of China (41361045, 41130531).

\*\* 通讯联系人, E-mail: zhangfei3s@163.com

Corresponding author, E-mail: zhangfei3s@163.com

quasi-synchronous Landsat OLI data to select sensitive band. We then used geo-statistics method, multiple linear regression model and SVM model to estimate the TDS of surface water, and compared with field measured TDS. The results show that SVM is the best model for TDS estimation and its coefficient  $R^2$  is 0.97, RMSE is 50.59. The accuracy of multiple linear regression model is very close to the SVM model, and the coefficient  $R^2$  is 0.9, RMSE is 66.55. The accuracy of Geo-statistical kriging interpolation method is the lowest, and coefficient  $R^2$  is 0.87, RMSE is 95.73. The SVM model prediction can better reflect the general characteristics in large area. The estimated results of TDS are in conformity with the actual field TDS data in Ebinur lake. Therefore, the SVM model has a great potential application for water quality prediction.

**Keywords:** TDS, SVM model, multiple linear regression model, geostatistics, Ebinur Lake watershed.

在西北极端干旱地区,绿洲是人类生存唯一的依托之地,它对水资源有着极强的依赖性,水成为了绿洲社会经济发展的最大制约因子<sup>[1-4]</sup>.水更是绿洲的依存基础<sup>[5]</sup>、是维持绿洲及荒漠-绿洲过渡带天然植被正常生长的重要因素,合理开采地表水资源,是维持绿洲社会经济的可持续发展的基本前提.水中的溶解性总固体 TDS (Total dissolved solids),其组分主要为溶解性的盐类.当水中含有大量的溶解性固体时,可能使水中带有苦咸味,因此 TDS 是水质评价的一项重要指标.地表水溶解性总固体(TDS)是地表水各组分浓度的总指标,是地表水水化学特性长期演变的最终结果,也是表征水文地球化学作用过程的重要参数<sup>[6-8]</sup>,地表水 TDS 的高低直接影响地表水的含盐量<sup>[9]</sup>.

目前,随着水质遥感技术的不断发展,可用遥感监测的水质指标逐渐增多.水质遥感估测模型的构建方法日益增多,自 2000 年以后,国内外的专家和学者在该研究领域常用的遥感估测主要方法有:线性回归、多元线性回归、聚类分析、主成分分析、BP (Back propagation) 神经网络和支持向量机等<sup>[10]</sup>,其中多元线性回归、BP 神经网络模型和支持向量机模型应用广泛,且估算和估测效果较好<sup>[11]</sup>.而地统计学是一种空间分析方法,它能有效地解释属性变量在空间上的分布、变异和相关特征<sup>[12]</sup>.该方法已经得到了良好的应用<sup>[13-18]</sup>;如张晓庆等<sup>[19]</sup>利用克里格插值法对扎龙湿地叶绿素 a 浓度进行了空间插值处理达到了理想精度,王水献等<sup>[20]</sup>采用地统计学经典理论对焉耆盆地不同时期地下水 TDS 和埋深在时间和空间的变异特征进行了分析.由于水质参数的空间变异特征既具有随机性又具有结构性,因此,可以应用地统计学方法和遥感估测模型进行预测,以反映这种空间的变异特征.

艾比湖流域属于国家“一带一路战略”的核心区,是全球变化敏感地区和生态系统典型脆弱地区.因此准确地评估艾比湖流域的地表水 TDS,对绿洲水资源利用和管理及其相关的生态环境的保护与地区经济建设都具有重要意义.本研究选取艾比湖流域野外实测地表水 TDS 数据,结合同步的 Landsat OLI 数据,开展基于遥感估测和克里格插值的艾比湖流域地表水 TDS 估算及其精度对比,评价两种估测方法适用性,选择更优估测模型,以期对艾比湖流域水体中 TDS 的监测、预测和评价提供参考.

## 1 实验部分 (Experimental section)

### 1.1 研究区域概况

艾比湖流域位于古尔班通古特沙漠西南缘,东经  $79^{\circ}50'$ — $85^{\circ}10'$ ,北纬  $43^{\circ}38'$ — $45^{\circ}52'$ ,该区域位于“丝绸之路经济带”上,是亚欧大陆的交通要道.常年来在气候条件和人类活动的共同影响下,土壤盐渍化、水土流失严重.流域内西、南、北三面环山,中间为谷地平原,东部有尾间湖,流域面积  $56021 \text{ km}^2$ .流域地处亚欧大陆腹地,远-离海洋,降水稀少,蒸发量大,光热充足,属典型的温带大陆性气候.艾比湖流域包括东南西北 4 个河区:东边为奎屯河区,南边为精河区,西边为博尔塔拉河区,北边为托里小河区.近年来随着艾比湖流域经济的发展,艾比湖湿地的水环境质量受到严重影响<sup>[21-22]</sup>.

### 1.2 数据源与数据处理

#### 1.2.1 水样信息的采集

本文中所使用的地表水 TDS 数据由 2015 年 5 月 19—5 月 28 日艾比湖流域野外调查所采集,共

53 个样品,采样路线按照沿河采样:大河沿子河(采样点 P10、P11);精河沿线(采样点 P12、P13、P14、P15、P16、P17、P18、P19、P20);博河沿线(采样点 P28、P29、P30、P31、P32、P33、P34、P35、P36、P37、P38、P39、P40、P41、P42、P43、P44、P45);奎屯河及其南部小河沿线干支流(采样点 P7、P8、P9、P46、P47、P48、P49、P50、P51、P52、P53);阿克奇苏河沿线(采样点 P1、P2、P3、P4、P5、P6);艾比湖周边小河沿线(采样点 P21、P22、P23、P24、P25、P26、P27),采样点如图 1 所示。

研究小组通过对艾比湖整个流域的实地勘察,采用河流排入区域加密等原则,采样点设置在湖水(水库)主要出入口、中心区、沿水流方向、滞留区等区域,河流的交汇处,农田周边等具有代表性的特殊区域。在水样采集时,使用便携式 GPS 导航仪对采样点进行经纬度测定,用人工水质采集仪器在水层表面 0.3—0.5 cm 处采集水样。在采集过程中将聚乙烯塑料瓶用要采集的水样进行同化,将聚乙烯塑料瓶拧紧;最后将样品进行编号。将采集的水样装箱,冷藏保存,尽快运回实验室进行数据测定<sup>[22]</sup>。本研究使用的 TDS 数据是德国 WTW (Wissenschaftlich Technische Werkstätten) 公司生产的 inoLab® Cond 7310 台式电导率测试仪所测。

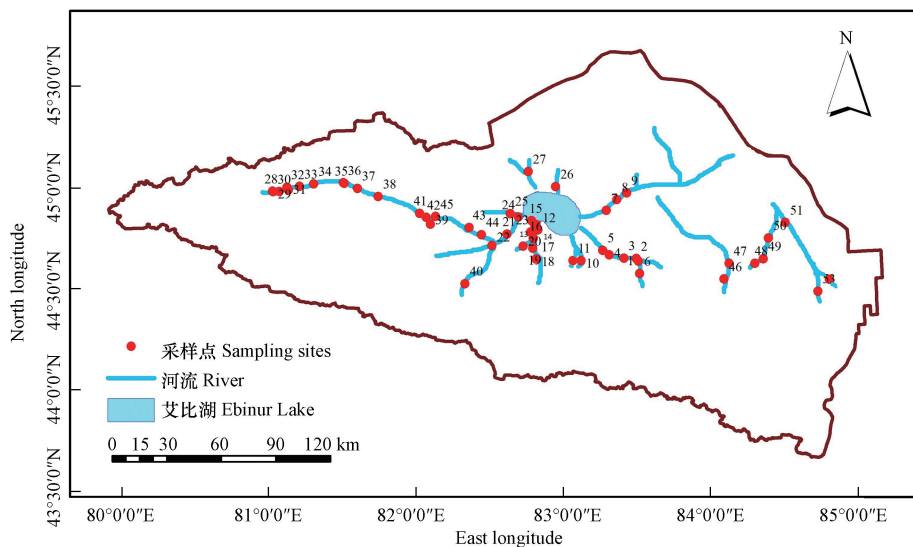


图 1 水体采样点示意图

Fig.1 Map of water sampling sites

### 1.2.2 遥感数据影像的处理

本研究选择与野外采样时间准同步的三景 Landsat OLI 数据(表 1),分别用于估测模型的构建和验证,轨道号分别为 145-29、146-29、147-29。对获取的影像进行辐射定标和大气校正计算,辐射定标的目的是将遥感影像的 DN (Digital number) 值转换为辐亮度;大气校正就是为了消除大气吸收和散射的影响,本研究采用 FLAASH 大气校正模型对影像进行大气校正,将辐亮度转换为地物表面反射率,FLAASH 大气校正是目前发展较为完善的大气校正方法之一,很多研究表明该模型的校正结果比其他模型的高,且计算时间快<sup>[23-24]</sup>。

## 1.3 研究方法

### 1.3.1 地表水 TDS 的诊断光谱的选取

各种地物具有丰富且独特的波谱信息,进行光谱诊断的目的,就是要从丰富的光谱信息中提取最能反映样本间差异的波段。本研究所用的诊断指数,主要用于研究 Landsat OLI 多光谱数据各个波段对地表水 TDS 的敏感程度的差异。诊断指数  $P_i$  的计算公式如下<sup>[25]</sup>:

$$P_i = A_i \times R_i \quad (1)$$

式中,  $A_i$  为样本光谱曲线在波段  $i$  上反射率值的标准差,  $R_i$  为波段  $i$  上的反射率数值与地表水 TDS 含之之间的相关系数,且诊断指数最大时的波段即为地表水 TDS 的诊断和估测波段。

表 1 野外采样时间、样点与准同步影像

Table 1 Sampling time, sampling sites and quasi-synchronous images

影像时间(轨道号) Images time (orbit)	野外采样时间 Sampling time	采样点 Sampling sites
2015-5-19(145-29)	2015-5-24、2015-5-26	P28、P29、P30、P31、P32、P33、P34、P35、P36、P37、P38
2015-5-26(146-29)	2015-5-19、2015-5-21	P1、P2、P3、P4、P5、P6、P7、P8、P9、P10、P11、P12、P13、P14、P15、P16、P17、P18、P19、P20、P21、P23、P24、P25、P26、P27、P39、P40、P41、P42、P43、P44、P45
2015-5-28(147-29)	2015-5-28	P46、P47、P48、P49、P50、P51、P52、P53

### 1.3.2 多元线性模型

由以上光谱分析选取研究区的地表水 TDS 反射率的敏感波段.为了将两者的关系进一步量化,尝试建立了两者之间的统计模型.根据“样本中位数检验法”<sup>[26]</sup>,从 53 个样本中剔除异常采样点以及仪器或人为操作而造成的异常值 6 个,将剩下的 47 个样本随机分成两组:一组 38 个采样点的地表水 TDS,用于建立预测模型;另一组 9 个地表水样本,作为检验样本用于验证建立的模型.

### 1.3.3 SVM 模型

支持向量机模型(SVM)能够在小样本学习的基础上对其他样本进行快速、准确的拟合预测,SVM 具有很好的泛化能力,能较好地解决小样本、非线性、高维数和局部极值等实际问题<sup>[27]</sup>.故本研究根据研究区实际情况,以及目前遥感估测水质的研究现状,采用 SVM 模型对艾比湖流域地表水 TDS 进行遥感估算.同理,将研究区的地表水水体反射率数据敏感波段反射率数据作为模型的输入值,地表水 TDS 数据为输出值建立艾比湖流域地表水 TDS 的 SVM 模型.

### 1.3.4 地统计方法

地统计学以区域化变量理论为基础,以半变异函数为基本工具,研究在空间分布上既具有随机性又具有结构性的自然现象<sup>[28]</sup>.本研究在地统计方法中选择了克里格空间插值法(Kriging),Kriging 插值方法着重于权重系数的确定,从而使内插函数处于最佳状态,即对给定点上的变量值提供最好的线性无偏估计<sup>[29]</sup>.其核心技术用半变异函数模型代表空间中随距离变化的函数,以无偏估计与最小估计变异函数的条件下,决定各采样点的权重系数,最后以各采样点与已求得的权重线性组合,来求空间任意点或块的内插估计值,克里格法的适用条件是区域化变量存在空间相关性<sup>[30]</sup>.

设  $Z(x)$  为区域化随机变量, $h$  为两样本点空间分隔距离, $Z(x_i)$  和  $Z(x_i+h)$  分别是区域化变量  $Z(x)$  在空间位置  $x_i$  和  $x_i+h$  处的实测值,那么变异函数  $r(h)$  由下式<sup>[30]</sup>进行估计:

$$y(h) = \frac{1}{2} \sum_{i=1}^{N(h)} [Z(x_i) - \bar{Z}(x_i+h)]^2 \quad i = 1, 2, 3, \dots, N(h) \quad (2)$$

式中, $N(h)$  为相距  $h$  的数据对数; $h$  是距离矢量; $y(h)$  仅仅是任意两点之间距离  $h$  的函数.常用的理论变异函数模型有:指数模型、球状模型、高斯模型、幂函数模型等<sup>[30]</sup>.指数模型一般形式为:

$$y(h) = \begin{cases} 0 & h = 0 \\ C_0 + C(1 - e^{-\frac{h}{a}}) & h > a \end{cases} \quad (3)$$

式中, $C_0$  是块金值; $(C_0+C)$  是基台值; $a$  是变程; $y(h)$  只有在最大间隔距离的 1/2 以内才有意义.

## 1.4 遥感估测模型的建立与验证方法

遥感定量估测模型的稳定性用拟合决定系数  $R^2$  来检验.拟合决定系数  $R^2$  进行的检验称为拟合优度检验, $R^2$  越大,模型的拟合度越高.模型的预测能力用均方根差(RMSE,来自建模与检验样本)来检验,RMSE 越小,模拟值(Predicted)和测量值(Measured)之间的偏差越小,预测能力越强.为了检验建立和优选出的估测模型的精度,采取随机预留的 9 个水体样品数据进行模型的精度验证,具体采用常用精度评价指标相对误差(RE)验证指标来进行检验,当 RE 的值越小,则估测模型的精度越高.

总均方误差(RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$



相对误差(RE):

$$RE = \sqrt{\frac{\sum (y_i - \bar{y})^2}{y_i}} \quad (5)$$

式中,  $y_i$  为采样点的水体 TDS, 即实测值,  $\hat{y}_i$  为利用回归模型求得的预测值,  $\bar{y}$  为样本标准差,  $k$  是自变量的个数,  $n$  为样本数目.

## 2 结果与讨论(Results and discussion)

### 2.1 艾比湖流域地表水 TDS 光谱反射特征研究及敏感波段选取

#### 2.1.1 艾比湖流域地表水 TDS 光谱反射特征研究

地物光谱特征是遥感手段鉴别地物的基本方法之一, 本研究从 Landsat OLI 数据上提取 53 个采样点水体光谱反射率曲线, 如图 2 所示. 从图 2 中可以看出, 采样点反射率曲线完全符合水体光谱反射率曲线特征, 因为 Landsat OLI 数据的分辨率为 30 m, 导致水体采样点可能存在混合像元问题, 光谱反射率曲线完全可以排除混合像元, 从图 2 中的 53 条水体光谱曲线可以判断, 本研究采样点均处于水体像元上, 因此基于遥感方法进行水质估算是可行的.

地表水光谱吸收特征与地表水 TDS 之间的物理关系或半经验模型是能否通过遥感手段进行 TDS 监测的前提, 本研究基于 Landsat OLI 数据的光谱特征, 对地表水水体光谱吸收特征与 TDS 之间的关系进行了探索性的研究, 选择不同含量的 TDS 及对应所在像元的光谱曲线如图 3 所示, 图中采样点 13(P13)的 TDS 为  $4230 \text{ mg} \cdot \text{L}^{-1}$ , 采样点 8 (P8)的 TDS 为  $2751 \text{ mg} \cdot \text{L}^{-1}$ , 采样点 50 (P50)的 TDS 为  $1505 \text{ mg} \cdot \text{L}^{-1}$ , 采样点 46 (P46)的 TDS 为  $577 \text{ mg} \cdot \text{L}^{-1}$ , 采样点 33 (P33)的 TDS 为  $95 \text{ mg} \cdot \text{L}^{-1}$ . 从图 3 可以看出, 在 430—1050 nm 时, 不能通过此段的水体反射率曲线来区分 TDS 的含量. 当在 1050—2450 nm 时, 光谱反射率值可以反映 TDS 的大小. 因此, 地表水 TDS 的差异可以通过水体反射率来表征, 进一步从光谱特征来判断发现 Landsat OLI 数据的第 4 (630—680 nm)、第 5 (840—885 nm)、第 6 (1560—1660 nm)、第 7 (2100—2300 nm) 波段为水体 TDS 的敏感性波段. 所以, 可以用以上波段表征地表水 TDS 的差异.

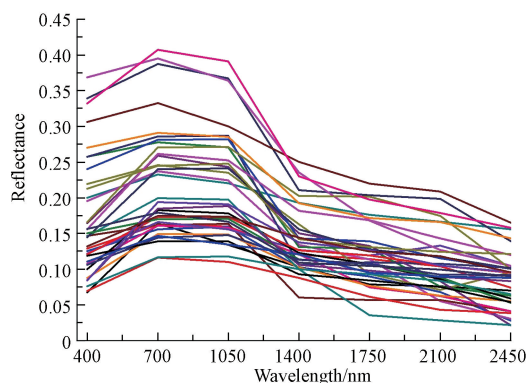


图 2 艾比湖流域水体光谱曲线(53 条光谱曲线)

Fig.2 The water spectrum curve in the Ebinur Lake watershed(Article 53 the spectral curve)

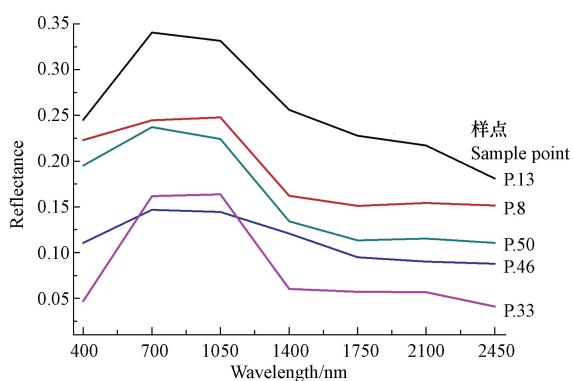


图 3 不同 TDS 水体光谱曲线

Fig.3 Water spectrum curves with different TDS contents

#### 2.1.2 艾比湖流域地表水 TDS 诊断光谱的选取

为了进一步准确、合理地提取敏感性波段, 利用光谱诊断指数进行敏感波段的选取, 将 Landsat OLI 遥感影像中对应野外 53 个采样点 7 个波段的多光谱反射率数据与地表水 TDS 数据进行相关性分析, 发现两者之间属于低度相关,  $0.3 < R < 0.5$ . 但是对反射率进行数学变换(倒数、对数、指数)后, 发现第 5、6、7 的 3 个波段反射率的倒数与水体 TDS 相关性  $R > 0.7$ , 属于高度相关. 且第 5、6、7 波段与 TDS 的相关性明显高于前 4 个波段, 因此, 地表水 TDS 在近红外波段和短波红外波段与光谱反射率呈显著相关. 分别计

算出 Landsat OLI 遥感影像 1—7 波段反射率倒数的  $R_i$ 、 $A_i$ 、 $P_i$ , 结果见表 2. 第 4、5、6 波段  $P_i$  项指标数值高于其他 4 个波段, 对地表水的 TDS 较为敏感, 该波段范围较适用于估测研究区地表水 TDS. 由于诊断指数将  $A_i$  和  $R_i$  结合在一起考虑, 用这种方法计算出的诊断波段将更为合理. 本研究通过两种方法选取的有效波段均为第 4、5、6 波段, 因此, Landsat OLI 遥感监测水体 TDS 的敏感性波段为 4、5、6 波段.

表 2 各波段与地表水 TDS 的相关系数、标准差以及诊断指数

Table 2 The correlation coefficient, standard deviation and diagnosis index of each band and surface water TDS

	波段 (Band)						
	B1	B2	B3	B4	B5	B6	B7
$R_i$ 相关系数	0.323	0.201	0.203	0.77	0.702	0.695	0.755
$A_i$ 标准差	4.3738	5.1624	3.2156	3.4311	2.9721	28.2068	2.16
$P_i$ 诊断指数	1.4127	1.037	0.6527	2.6419	2.0864	19.6037	1.6308

注: B1、B2、B3、B4、B5、B6、B7 代表 Landsat OLI 影像的 7 个波段.

Note: B1, B2, B3, B4, B5, B6 and B7 representative the Landsat OLI image of 7 bands.

## 2.2 艾比湖流域地表水 TDS 遥感估算模型的建立

### 2.2.1 艾比湖流域地表水 TDS 多元线性估算模型

本研究利用 SPSS 19.0, 将已选好的 4、5、6 等 3 个敏感波段反射率的倒数作为自变量, 使用多元线性回归模型进行地表水 TDS 的遥感估测. 建立的多元线性估测模型为:

$$\hat{Y} = -381 + 768 \frac{1}{B_4} - 1350 \frac{1}{B_5} + 705 \frac{1}{B_6} \quad (6)$$

该模型数据的决定系数  $R^2$  为 0.605, RMSE 为 337.93; 多元线性估算模型为传统的估测模型, 模型估测精度较高.

### 2.2.2 艾比湖流域地表水 TDS 遥感 SVM 估算模型

本研究旨在寻求干旱区艾比湖流域地表水 TDS 的最优估算模型, 因此为进一步选取遥感估测 SVM 模型的非线性结构建立艾比湖流域地表水 TDS 估测模型. 通过 MATLAB 2014a 作为软件平台, Landsat OLI 4、5、6 的 3 个波段的反射率的倒数为自变量, TDS 为因变量, 使用 SVM 模型进行地表水 TDS 的遥感估测. 并结合模型的拟合精度采用支持向量机回归 (SVM) 算法实现建模<sup>[31]</sup>. 利用 SVM 算法构建估测模型时, 核函数类型、惩罚系数  $C$  与核函数参数  $g$  的选取对结果影响较大, 本文利用文献[32]所介绍的方法, 选用 RBF 核函数, 通过交叉验证确定. 为降低搜索难度与节省计算时间, 交叉验证将采用分步网格搜索法 (grid-search) 进行, 即第一步确定参数大概的取值范围, 第二步则在第一步所确定的范围内进行搜索以获取最佳参数取值.

本研究建立遥感估测 SVM 模型的决定系数  $R^2$  为 0.667, RMSE 为 169.82; 由该模型得到的地表水 TDS 预测值与实测值的关系见图 4.

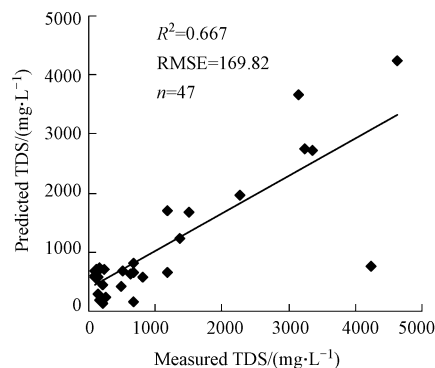


图 4 TDS 遥感估测 SVM 模型预测值与实测值的关系

Fig.4 The relationship between observed and predicted TDS based on remote sensing inversion SVM model

### 2.3 艾比湖流域地表水 TDS 的克里格空间插值

为了进一步有效地解释艾比湖流域地表水 TDS 在空间上的分布、变异和相关特征并建立空间预测模型,利用地统计方法建立艾比湖流域地表水 TDS 估测模型.经过对艾比湖流域实测的 QQplot 图,发现其空间差异比较大,变化比较剧烈.因此,艾比湖流域实测 TDS 数据很接近正态分布.所以,本研究数据符合克里格插值预测要求,利用地统计软件 GS+V 9.0,计算地表水水体中的 TDS 半方差函数,分别经过线性、球状、指数和高斯模型模拟,以残差平方和最小、决定系数最大为依据,选择出的最佳模拟模型为球状模型,其相关参数分别为块金值  $C_0 = 0.0116$ , 基台值  $(C_0 + C) = 0.125$ , 结构比  $C / (C_0 + C) = 0.072$ , 变程  $a = 14142 \text{ m}$ ,  $R^2 = 0.513$ , 残差平方和为 1916.

### 2.4 艾比湖流域地表水 SVM 模型、多元线性模型与空间插值的精度对比

TDS 预测值与实测 TDS 的关系见图 5.由克里格预测模型得到的 9 个预测点 TDS 预测值与实测 TDS 的关系见图 5.从验证实测值与预测值的散点图 5 中可以看出,克里格法预测值与真实值间的  $R^2$  为 0.87, RMSE 为 95.73.线性回归模型预测值与真实值间的  $R^2$  为 0.90, RMSE 为 66.55.从 SVM 模型的验证实测值与预测值的散点图中可以看出,预测值与真实值间的  $R^2$  为 0.97, 遥感估测 SVM 模型预测值的均方根误差为 50.59.因此,遥感估测 SVM 模型的预测精度较高于传统的地统计方法克里格插值法和多元线性回归模型.

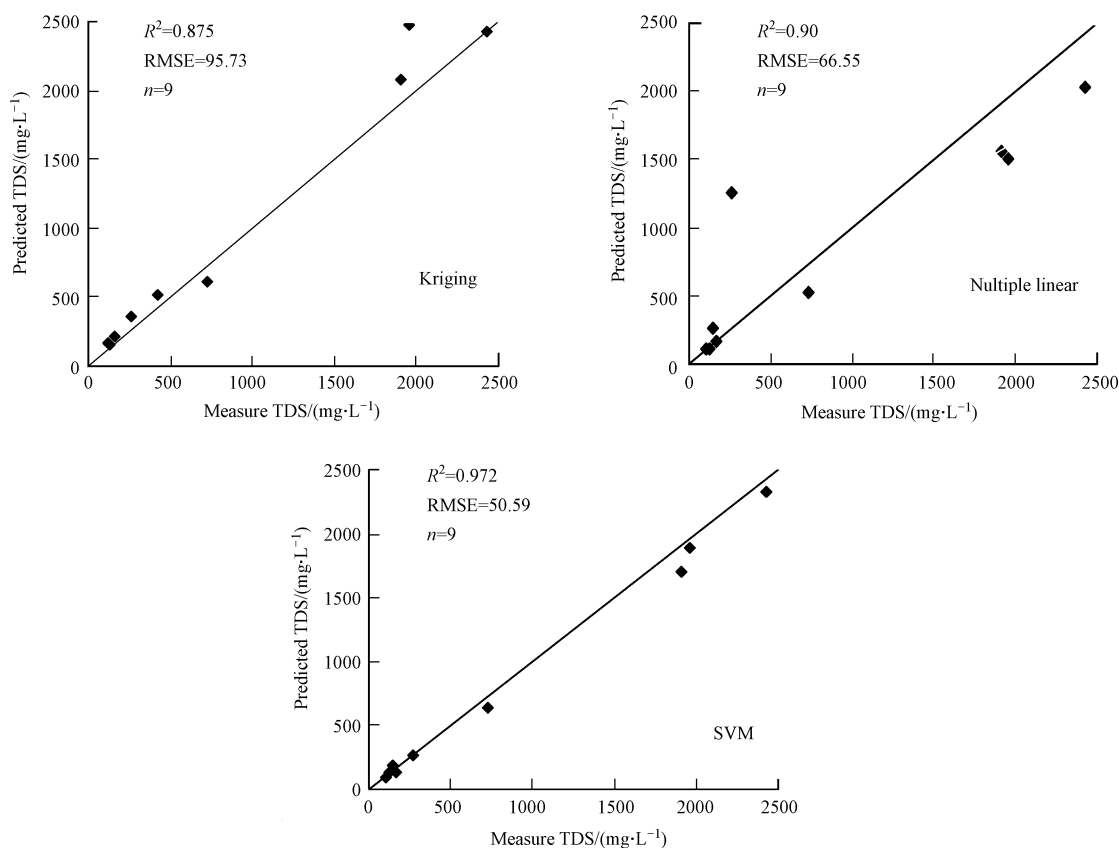


图 5 克里格插值模型、多元线性模型和 SVM 模型精度对比

Fig.5 Comparison of accuracy between of multiple linear regression, SVM and Kriging model

为进一步对遥感估测多元线性模型、SVM 模型和克里格空间插值法进行验证.由表 3 可以看出,在多元线性回归模型预测结果中,采样点 48(P48)的预测 TDS 的相对误差较大,为 27.6%;采样点 6(P6)和采样点 16(P16)的预测 TDS 的相对误差较小,均为 8.9%,其余采样点预测值的相对误差居其间;在遥感估测克里格插值法预测值中,采样点 21(P121)和采样点 42(P42)的预测值的相对误差较大,分别为 33.74%和 36.60%;采样点 1(P1)和采样点 26(P26)的预测值的相对误差较小,为 9.07%和 9.1%,其余采样点的预测值的相对误差居其间.在遥感估测 SVM 模型预测值中,采样点 16(P16)和采样点

42(P42)的预测值的相对误差较小,分别为 2.35% 和 2.44%;34 号采样点的预测值的相对误差较大,为 23.96%,其余采样点的预测值的相对误差居其之间.相比较多元线性回归模型、SVM 模型和克里格插值法对艾比湖流域地表水的预测精度,发现克里格插值法的验证精度低于 SVM 模型和多元线性回归模型,SVM 模型和多元线性回归模型精度比较接近,但 SVM 模型精度稍优于多元线性模型,进一步说明地表水水质估算模型中遥感估算模型精度高于传统的统计模型.

表 3 TDS 的遥感估测 SVM 模型和克里格插值预测值及精度

Table 3 Prediction of TDS and their accuracies by remote sensing inversion SVM model and Kriging spatial interpolation method

采样点 Sampling point	实测值 Measured value	多元线性模型 Multiple linear model		克里格插值模型 Kriging model		SVM 模型 SVM model	
		预测 Predicted/ ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 Relative error/%	预测 Predicted/ ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 Relative error/%	预测值 Predicted/ ( $\text{mg}\cdot\text{L}^{-1}$ )	相对误差 Relative error/%
P1	2425	2031	-16	2205	-9.07	2321	4.28
P6	123	112	-8.9	143	16.2	136	10.98
P12	163	198	21	218	33.74	182	12.02
P16	428	466	8.9	512	19.62	438	2.35
P26	1910	1500	-21	2084	9.10	1891	-0.98
P34	127	108	-14.9	162	27.55	96	-23.96
P42	265	227	14.3	362	36.60	271	2.44
P48	730	528	27.6	613	-16.02	632	-13.34
P52	1960	1566	20	2475	26.27	1708	-12.85

## 2.5 艾比湖流域地表水 TDS 的空间分布

在艾比湖流域,TDS 的分布如图 6 所示,以艾比湖为中心,最大值出现在艾比湖最近的入湖河流中,呈现出中心含量高,四周低的特点.在入湖河流中,西部绿洲内的博尔塔拉河水体 TDS 比东部绿洲内阿其克苏河和奎屯河水体的 TDS 低,主要是因为阿其克苏河和奎屯河流经的区域受土壤盐渍化的影响较为严重,水体的含盐量较高,致使 TDS 上升,由此可见,艾比湖流域 TDS 主要受土地利用的影响;盐渍土是流域 TDS 的主要贡献者.但对于河流的出山口,冰雪融水的汇入,致使其河流在上游地区 TDS 较小.

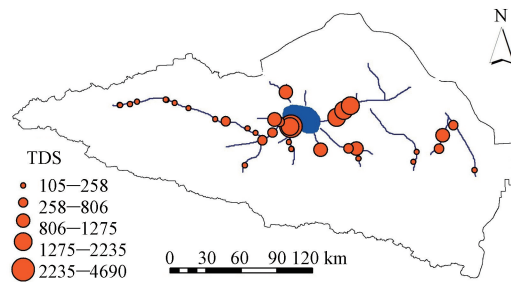


图 6 TDS 空间分布图

Fig.6 Spatial distribution of TDS

为了进一步比较 TDS 在流域内的分布异同,本研究利用软件 ArcGIS10.2 对艾比湖流域水体 TDS 进行插值<sup>[33]</sup>,为了突显插值效果,本研究对河流分别做 100 m、500 m、1000 m 和 1500 m 缓冲区并进行插值,结果发现 1000 m 缓冲区的差值图效果比较好,能够准确地表达艾比湖流域的 TDS 分布情况,TDS 在整个流域内的空间分布如图 7 所示(A 图为遥感多元线性回归模型插值图,B 图为克里格预测插值图,C 图为遥感 SVM 模型插值图).由(图 7A、B、C)可知 TDS 的空间分布格局大体与盛行风风向一致并有向湖中心集中的趋势.TDS 高值区主要分布在艾比湖北部及其流域东北部的小河区,艾比湖的东北部河流交汇处地表水 TDS 高于周边河段,因为东北部地区地表水 TDS 比较高.总体而言,高值都是以艾比湖为中心在研究区的东北方向分布,形成整个流域中间值高,两边值较低的空间分布格局.



图 7(A、B、C)的渐变趋势是两种预测方法本质的体现,图中黑色虚线框中为流域主要盐渍区域,该区域的阿克其苏河和托托河在中下游 TDS 偏高,而克里格插值图,图 7(A)中的 TDS 值相对偏低,因此克里格插值图存在误差,因为地统计克里格空间插值是基于区域化变量理论,通过空间相关的随机函数模型计算可获取变量的线性加权组合,再估算采样点的地表水 TDS.而遥感估测模型的预测值可以反映艾比湖流域地表水 TDS 的总体空间分布特征及不同含量处的空间变异特性,如图 7A、B 所示,特别是博河河段遥感预测的 TDS 渐变趋势更加接近实测值,这主要是由于遥感估测模型是以地表水的光谱反射率为自变量,根据遥感估测模型计算 TDS,这也是地表水光谱反射率空间异质性的反映.

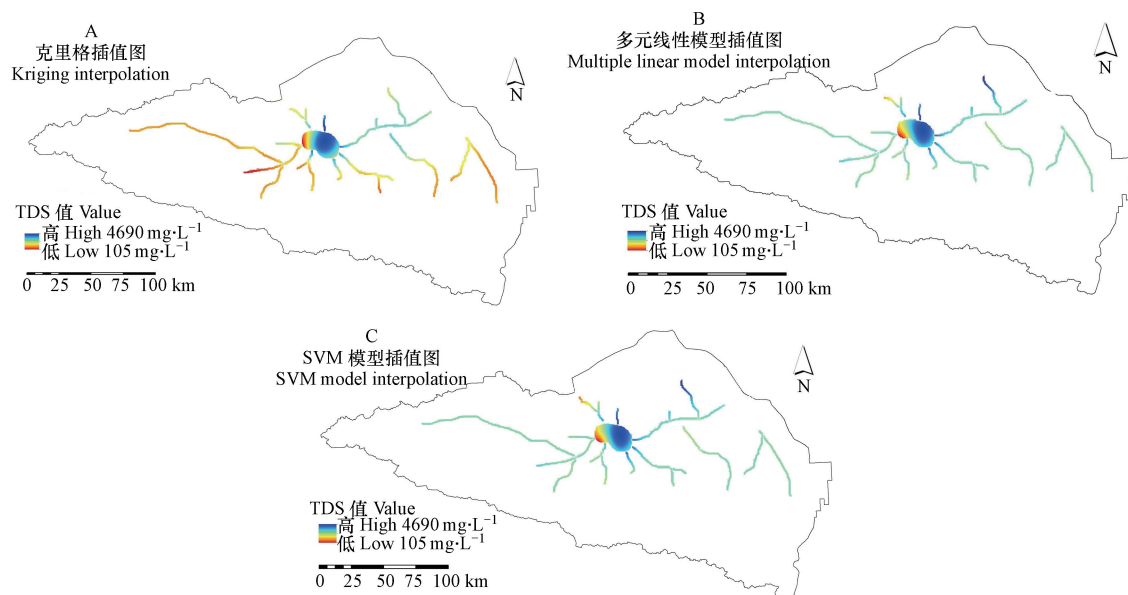


图 7 遥感多元线性回归模型(A)、克里格预测(B)、遥感 SVM 模型(C)插值图

Fig.7 The interpolation figure of Multiple linear regression model (A), Kriging prediction(B), SVM model (C)

### 3 结论 (Conclusion)

以艾比湖流域为研究对象,在野外调查采样和分析的基础上,利用地表水 TDS 信息与 Landsat OLI 多光谱数据基于多元线性回归模型、SVM 模型定量估测地表水 TDS,并将其结果与克里格差值预测进行对比,得出了以下结论:

(1) 利用多元线性回归模型、SVM 模型和克里格插值法建立了地表水 TDS 和 Landsat OLI 数据之间的定量关系,多元线性回归模型的拟合决定系数  $R^2$  为 0.605,均方根误差 RMSE 为 337.93,克里格插值法的地表水 TDS 插值模型拟合决定系数  $R^2$  为 0.513,SVM 估测模型的拟合决定系数  $R^2$  为 0.667, RMSE 为 169.82.

(2) 对构建的地表水 TDS 多元线性回归模型、SVM 模型和克里格插值法模型进行精度对比发现,SVM 模型为最优估测模型,拟合决定系数  $R^2$  为 0.97, RMSE 为 50.59,而多元线性回归模型的精度稍低于 SVM 模型,多元线性模型的拟合决定系数  $R^2$  为 0.9, RMSE 为 66.55.克里格插值法预测精度最低,拟合决定系数  $R^2$  为 0.87, RMSE 为 95.73.说明遥感估测 SVM 模型的建立对地表水 TDS 的预测精度提高较为关键.

(3) 通过构建地表水 TDS 与 Landsat OLI 数据之间的关系的遥感估测 SVM 模型以及用实测数据对模型进行验证,发现遥感估测 SVM 模型结果近似可以代替实测结果.

通过分析研究,遥感估测模型 SVM 预测值在大区域能较好地反映出艾比湖流域 TDS 的总体特征.其预测结果也与艾比湖流域水体 TDS 的实际分布相吻合.遥感估测 SVM 模型不仅可以更方便和快捷地监测水体信息和水质的变化情况,而且对干旱区水污染遥感监测也具有一定的参考意义.

本研究中,由于受到当地自然条件的限制,样点的选择不是均匀的,样点间的间距也有大有小,这可

能也是导致地统计学方法的预测效果比较差一个根本原因.假如能够忽略自然条件的影响,进行等间距采样,那么预测的效果可能要优于现在的结果,但是效果能不能达到遥感定量预测还不得而知.因为这样做数据量、工作量都十分的巨大,目前并没有尝试,因此,在以后的工作当中作者将会进一步研究这个问题.

该研究方法是否适用于其他地区还有待进一步验证;其他数据源是否会达到同样的效果还需进一步深入的研究和考证;本研究最终的结论是遥感估测 SVM 模型优于多元线性回归和克里格插值模型,虽然遥感估测 SVM 模型的非线性结构能够较为真实、准确地估测出 TDS 在艾比湖流域的浓度分布.但在本研究中仅仅是 SVM 模型在构建水质遥感估测模型中的一个初步尝试,为了进一步提高 SVM 在该领域内的应用能力,将从模型输入数据的筛选、核函数的选择以及各参数的优化方面开展深入的研究.

#### 参考文献 (References)

- [ 1 ] 马金珠,李吉均.塔里木盆地南缘人类活动干扰地下水变化及其生态环境效应[J].自然资源学报,2001,16(2):134-139.  
MA J Z,LI J J. Impact of human activities on groundwater and the effect on eco-environment in southern Tarim Basin[J]. Journal of Natural Resources, 2001, 16(2):134-139(in Chinese).
- [ 2 ] 刘昌明,陈志恺.中国水资源现状评价和供需发展趋势分析[M].北京:中国水利水电出版社,2001.  
LIU C M,CHEN Z K. Present evaluation of China water resources and development trend analysis of supply and demand situation[M]. Beijing: China Water Power Press,2001(in Chinese).
- [ 3 ] COLLIN M L,MELLOUL A J.Combined land-use and environmental factors for sustainable groundwater management [J].Urban Water, 2001, 3(3):229-237.
- [ 4 ] 王雪蕾,王新新,朱利,等.巢湖流域氮磷面源污染与水华空间分布遥感解析[J].中国环境科学,2015,35(5):1511-1519.  
WANG X L, WANG X X, ZHU L, et al.Spatial analysis on diffuse pollution and algal bloom characteristic with remote sensing in Chao Lake Basin[J].China Environmental Science,2015,35(5):1511-1519(in Chinese).
- [ 5 ] 任建华.黑河流域水资源开发对生态环境的影响[J].水土保持通报,2005,25(4):92-96.  
REN J H.Effects of water resources exploitation on eco-environment of Heihe River basin[J].Bulletin of Soil and Water Conservation, 2005,25(4):92-96(in Chinese).
- [ 6 ] 章光新,邓伟,何岩.洮儿河流域地下水 TDS 时空变异特征研究[J].水土保持学报,2005,19(2):122-124.  
ZHANG X G, DENG W, HE Y.Characteristics of spatial and temporal variation of TDS of Groundwater in Tao'er River basin[J].Journal of Soil and Water Conservation, 2005,19(2):122-124(in Chinese).
- [ 7 ] 于玲红,段浩洁,韩佩江,等.呼伦湖冰封期溶解性总固体的垂直分布特征[J].宁夏大学学报(自然科学版),2015,36(3)284-287.  
YU L H, DUAN H H, HAN P J, et al. The vertical distribution characters of total dissolved solids(TDS) in frozen period Hulun lake[J]. Journal of Ningxia Univerity (Natural Science Edition),2015,36(3)284-287(in Chinese).
- [ 8 ] 李东青,梁籍,张立燕.密云库区 1991-2011 年水质变化趋势研究[J].中国环境科学,2015,35(6):1675-1685.  
LI D Q, LIANG J, ZHANG L Y.The research of water quality trend in the Miyun Reservoir from 1991 to 2011[J].China Environmental Science, 2015,35(6):1675-1685(in Chinese).
- [ 9 ] SALAMA R B, OTTO C J, FITZPATRICK R W. Contributions of groundwater conditions to soil and water salinization[J]. Hydrogeology Journal, 1999, 7(1): 46-64.
- [ 10 ] 烟贯发,杜百利,张冬有,等.松花江哈尔滨段悬浮物含量遥感反演与克里格插值预测精度对比[J].湿地科学,2015,13(2):180-189.  
YAN G F, DU B L, ZHANG D Y, et al.Comparison of prediction accuracies of contents of suspended solids by remote sensing inversion and kriging spatial interpolation at harbin section of Songhua River[J].Wetland Science, 2015,13(2):180-189(in Chinese).
- [ 11 ] CAMPBELL G, PHINN S R, DEKKER A G, et al. Remote sensing of water quality in an Australian tropical freshwater impoundment using matrix inversion and MERIS images[J]. Remote Sensing of Environment, 2011, 115(9): 2402-2414.
- [ 12 ] 胡炳清,易鹏,段宁.影响精确界定我国酸雨区空间分布的因素探讨[J].中国环境科学,2015,35(3):917-924.  
HU B Q, YI P, DUAN N. Study on the influence factors of accurate definition of the spatial distribution of acid rain area in China[J]. China Environmental Science, 2015,35(3):917-924(in Chinese).
- [ 13 ] ZHANG X Y,SUI Y Y, ZHANG X D, et al.Spatial variability of nutrient properties in black soil of northeast China[J]. Pedosphere,2007, 17(1):19-29.
- [ 14 ] 竞霞,黄文江,王纪华,等.采用 IRS-P6 遥感数据监测密云水库水质及评价营养状况[J].农业工程学报,2008,24(增刊):13-17.  
JING X, HUANG W J, WANG J H, et al. Water quality monitoring and trophic state evaluation in Miyun reservoir based on IRS-P6 remote sensing data[J]. Transactions of the CSAE, 2008,24(Supp.2):13-17(in Chinese).
- [ 15 ] 马媛,师庆东,杨建军.干旱区典型流域土壤微量元素的空间变异特征研究[J].干旱区地理,2006,29(5):682-687.  
MA Y, SHI Q D, YANG J J.Spatial variability of characters of soil trace elements in a watershed of arid area[J].Arid Land Geography,

- 2006,29(5):682-687(in Chinese).
- [16] ZHANG C S, MCGRATH D. Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of southeastern Ireland from two different periods[J]. *Geoderma*, 2004, 119(3): 261-275.
- [17] GLENN N F, CARR J R. The use of geostatistics in relating soil moisture to RADARSAT-1 SAR data obtained over the Great Basin, Nevada, USA[J]. *Computers & Geosciences*, 2003, 29(5): 577-586.
- [18] LI L, LI J, CHENG G. The dynamic monitoring of erhai cyanobacteria based on landsat data[J]. *Journal of Water Resources Research*, 2016, 5(2): 167-173.
- [19] 张晓庆, 李苗, 臧淑英. 扎龙湿地叶绿素 a 浓度的协同克里格插值[J]. *中国农学通报*, 2013,29(8):160-164.  
ZHANG X Q, LI M, ZANG S Y. Spatial interpolation of the chlorophyll-a concentration in zhalong wetland based on Cokriging[J]. *Chinese Agricultural Science Bulletin*, 2013,29(8):160-164(in Chinese).
- [20] 王水献, 王云智, 董新光. 焉耆盆地浅层地下水埋深与 TDS 时空变异及水化学的演化特征[J]. *灌溉排水学报*, 2007, 26(5)90-93.  
WANG S X, WANG Y Z, DONG X G. The spatio-temporal variation of shallow groundwater TDS depth and its evolution characteristic of water chemistry in Yanqi Basin[J]. *Journal of Irrigation and Drainage*, 2007, 26(5): 90-93(in Chinese).
- [21] ZHANG F, TASHPOLAT T, VERNER C, et al. Evaluation of land desertification from 1990 to 2010 and its causes in Ebinur Lake region, Xinjiang China[J]. *Environ Earth Sci*, 2015, 73(9): 5731-5745.
- [22] 周梅, 张飞, 塔西甫拉提·特依拜, 等. 艾比湖流域水化学指标平面分布特征[J]. *环境化学*, 2015, 34(9): 1694-1702.  
ZHOU M, ZHANG F, TASHPOLAT TIYIP, et al. Spatial distribution of hydro-chemistry characteristics and water quality in Ebinur Lake Basin[J]. *Environmental Chemistry*, 2015, 34(9): 1694-1702(in Chinese).
- [23] ABDERRAZAK HARTI, RACHID LHISSOU, KAREM CHOKMANI, et al. Spatiotemporal monitoring of soil salinization in irrigated Tadla Plain(Morocco) using satellite spectral indices[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2016, 50: 64-73.
- [24] YANG Y H, LIU Y X, ZHOU M X, et al. Landsat 8 OLI image based terrestrial water extraction from heterogeneous backgrounds using a reflectance homogenization approach[J]. *Remote Sensing of Environment*, 2015, 171: 14-32.
- [25] 樊彦国, 李潭潭, 李祥昌. 基 Landsat8 的黄河三角洲盐渍化反演[J]. *山东农业科学*, 2015, 47(2): 119-124.  
FAN Y G, LI T T, LI X J. Soil salinization retrieval for the Yellow River delta based on Landsat 8[J]. *Shandong Agricultural Sciences*, 2015, 47(2): 119-124(in Chinese).
- [26] 吕恕, 朱宏. 统计数据中异常值的检验方法讨论[J]. *东北师大学报自然科学版*, 1993, 25(3): 27-31.  
LV S, ZHU H. A discussion about the test of outliers in statistical data[J]. *Journal of Northeast Normal University*, 1993, 25(3): 27-31(in Chinese).
- [27] 郭亚男, 吴泽宁, 高建菊. 基于主成分分析的支持向量机需水预测模型及其应用[J]. *中国农村水利水电*, 2012(7): 76-83.  
GUO Y N, WU Z Y, GAO J J. Application of support vector machines based on principal component analysis in water demand prediction [J]. *China's rural water conservancy and hydropower*, 2012(7): 76-83(in Chinese).
- [28] DING J L, YU D L. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments[J]. *Geoderma*, 2014, 235-236: 316-322.
- [29] 范晓梅, 刘高焕, 刘红光. 基于 Kriging 和 Cokriging 方法的黄河三角洲土壤盐渍化评价[J]. *资源科学*, 2014, 36(2): 321-327.  
FAN X M, LIU G H, LIU H G. Evaluating the spatial distribution of soil salinity in the Yellow River delta based on Kriging and Cokriging methods[J]. *Resources Science*, 2014, 36(2): 321-327(in Chinese).
- [30] 徐利岗, 周宏飞, 潘锋. 三工河流域山地-绿洲-荒漠系统降水空间变异性研究[J]. *地理学报*, 2016, 71(5): 731-742.  
XU L G, ZHOU H F, PAN F. Spatial variability of precipitation for mountain-oasis-desert system in the Sangong River Basin[J]. *Acta Geographica Sinica*, 2016, 71(5): 731-742(in Chinese).
- [31] GOTTSCHALK L, BATCHVAROVA B, GRYNING S B, et al. Scale aggregation comparison of flux estimates from XOPLEX[J]. *Agric for Meteorol*, 1999, 98: 103-119.
- [32] 林卉, 梁亮, 等. 基于支持向量机回归算法的叶面指数高光谱遥感反演[J]. *农业工程学报*, 2013, 29(11): 139-146.  
LIN H, LIANG L, et al. Wheat leaf area index inversion with hyperspectral remote sensing based on support vector regression algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(11): 139-146(in Chinese).
- [33] 刘双印, 徐龙琴, 李道亮, 等. 基于时间相似数据的支持向量机水质溶解氧在线预测[J]. *农业工程学报*, 2014, 30(3): 155-162.  
LIU S Y, XU L Q, LI D L, et al. Online prediction for dissolved oxygen of water quality based on support vector machine with time series similar data[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2014, 30(3): 155-162(in Chinese).