

基于 Bayes 判别理论的地下水化学分类的分析方法

李 彬^{1,2}, 史海滨¹, 李 祯¹, 张建国³

(1. 内蒙古农业大学 水利与土木建筑工程学院, 内蒙古 呼和浩特 010018;
2. 内蒙古农牧业科学院资源环境与检测技术研究所, 内蒙古 呼和浩特 010031; 3. 巴彦淖尔市水利科学研究所, 内蒙古 巴彦淖尔 015000)

摘 要: 将 Bayes 判别分析方法应用于地下水化学类型判别与分类中, 建立了地下水化学类型综合评判的 Bayes 判别分析模型。模型选用 $\text{Na}^+ + \text{K}^+$ 、 Ca^{2+} 、 Mg^{2+} 、 HCO_3^- 、 Cl^- 、 SO_4^{2-} 等 6 个指标作为判别因子; 将地下水化学类型分为 3 种, 作为 Bayes 判别分析的 3 个正态总体; 以内蒙古河套灌区地下水实测数据作为训练样本, 建立 Bayes 线性判别函数; 以 Bayes 线性判别函数计算待判样品的 Bayes 判别函数值, 以最大值对应的总体作为样品所归属的总体; 最后以刀切法对判别准则进行评价以检验模型的优良性。结果表明, Bayes 判别分析模型误判率低, 识别正确率达 82.5%, 输出结果正确率达 86.6%。与传统分类方法相比, Bayes 判别分析模型提供的判别分类结果, 具有更明晰、易懂的水化学类型信息。

关键词: 内蒙古河套灌区; 地下水化学类型; Bayes 判别; 分类
中图分类号: P641 **文献标志码:** A

An analyzing method for chemical classifications of groundwater based on the Bayes Discriminant Theory

LI Bin^{1,2}, SHI Hai-bin¹, LI Zhen¹, ZHANG Jian-guo³

(1. College of Water Conservancy and Civil Engineering, Inner Mongolia Agriculture University, Hohhot, Inner Mongolia 010018, China; 2. Institute of Resources and Environment and Testing Technology of Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot, Inner Mongolia 010031, China; 3. Institute of Water Conservancy Science, Inner Mongolia Bayannur, Bayannur, Inner Mongolia 015000, China)

Abstract: Bayes discriminant analysis method was applied into the discrimination and classification of groundwater chemical types to set up a Bayes discriminant analysis model for synthetic evaluation of groundwater chemical types. Six indexes including $\text{Na}^+ + \text{K}^+$ 、 Ca^{2+} 、 Mg^{2+} 、 HCO_3^- 、 Cl^- 、 SO_4^{2-} were selected as the differentiation parameters in this model. Chemicals in the groundwater could be classified into three types that were chosen as the three normal populations for Bayes discriminant analysis. Groundwater in Hetao irrigation district was utilized as the trial sample for data collection to set up the linear discriminant function of Bayes. The Bayesian linear discriminant function was then employed to calculate and determine the Bayes discriminant function values of the samples. The population with the maximal value was thereby used as the population for entire samples. Finally, in order to test the validity of this model, the cutting ring method was applied to evaluate the discriminate criterion. Results of this research showed that the misjudgment rate of this Bayes discriminant analysis model was low, reaching as high as 82.5% in correct recognition rate and 86.6% in output accuracy. Compared with traditional method, this Bayes discriminant analysis model could provide classification outcome that was clear and informative information for chemicals in water.

Keywords: Inner Mongolia Hetao Irrigation District; chemical type of groundwater; Bayes discriminant; classification

地下水的化学成分与其形成、赋存条件有着密切关系,是水-岩相互作用的产物^[1]。水中阴阳离子的组分含量,反映了水化学环境、水-岩相互作用、补排循环特征的差异^[2]。地下水的化学状况与农业生产、环境和生活质量有着密切的关系。盐渍化灌区的浅层地下水更有其特殊性,受原生地球化

收稿日期: 2014-06-12
基金项目: 国家“十二五”科技支撑计划项目(2011BAD29B03); 内蒙古科技攻关与政府专项科研项目(nsn200767)
作者简介: 李 彬(1982—), 男, 内蒙古集宁县人, 博士, 主要从事节水灌溉新技术和资源环境方向的研究。E-mail: libin-kim@sina.com。
通信作者: 史海滨(1961—), 男, 山西太谷人, 教授, 博士生导师, 主要从事节水灌溉原理及应用研究。E-mail: shi-haibin@sohu.com。

学环境和人类农业活动的共同影响,随着地下水化学成分的不断演化,其水化学类型也随之改变。水化学类型的改变可反映水环境的变化。作为灌区地下水质量评价的组成部分之一,进行水化学类型研究可为灌区地下水资源的合理利用和环境保护提供依据。

关于地下水化学分类的方法有很多,常用的方法包括:舒卡列夫分类法、布罗茨基分类法、阿廖金分类法、模糊聚类分析法、皮伯三线图解(Piper)法、多矩形图解(MRD)法^[3-7]。这些方法各有其优缺点,其中舒卡列夫分类法在地下水化学分类中的应用较为广泛,图解法也因为其不受人为因素影响的优点被广泛应用。

然而在利用上述方法进行地下水水化学类型分类时,计算较复杂,使用不方便,为此,本文尝试应用一种基于 Bayes 判别分析法的地下水化学类型分类模型,以达到在对地下水水化学类型分类时,计算方便,误判率低的目的。

1 Bayes 判别理论^[8-12]

1.1 判别分析的基本思想

简单地说,判别分析就是根据掌握的历史上每个类别的若干样本的数据信息,总结出客观事物分类的规律性,建立判别函数和判别准则,然后,当遇到新的样本点时只需根据总结出来的判别函数和判别准则就能够判别该样本点所属类别。而在地下水化学分类中,就是根据地下水样本中阴阳离子的信息,按照一定的分类方法,最终确定样本的类型。这使得将 Bayes 判别分析法应用于地下水化学分类成为可能。

用统计学的语言来描述 Bayes 判别分析就是:已知有 g 个 p 维总体 $G_1, G_2, \dots, G_g$, 每个总体 G_i , 可认为是属于 G_i 的指标 $X = (X_1, X_2, \dots, X_p)^T$ 取值的全体,它们分别具有互不相同的 p 维概率密度函数 $f_1(x), f_2(x), \dots, f_g(x)$, 对于任一给定的新样品关于指标 X 的观测值 $x = (x_1, x_2, \dots, x_p)^T$, 我们要判断该样品属于这 g 个总体中的哪一个。因而判别分析是根据所研究个体的某些指标的观测值来推断该个体所属类型的一种统计方法。

1.2 Bayes 判别函数

设两个 p 维正态总体 G_1, G_2 , 其概率密度函数为:

$$f_i(x) = (2\pi)^{-\frac{p}{2}} \left| \sum_i \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)^T \sum_i^{-1} (x - \mu_i) \right\} \quad (1)$$

式中: $\mu_i, \sum_i$ 为两总体的均值向量和协方差矩阵,

$|\sum_i|$ 为 $\sum_i$ 的行列式, $i = 1, 2$ 。

假设 $\sum_1 = \sum_2 = \sum$, 根据马氏距离和判别函数的概念, 可得

$$\frac{f_1(x)}{f_2(x)} = \exp \left\{ \frac{1}{2} \left[(x - \mu_2)^T \sum^{-1} (x - \mu_2) - (x - \mu_1)^T \sum^{-1} (x - \mu_1) \right] \right\} = \exp \{ W(x) \} \quad (2)$$

式中:

$$W(x) = \left[x - \frac{1}{2} (\mu_1 + \mu_2) \right]^T \sum^{-1} (\mu_1 - \mu_2) \quad (3)$$

称为 Bayes 判别函数。

1.3 多正态总体的 Bayes 判别

式(3)可方便地推广到多正态总体的判别分析中。设 g 个 p 维总体 $G_1, G_2, \dots, G_g$, 其概率密度函数为式(1), 此时, $i = 1, 2, \dots, g$ 。假设各正态总体的协方差矩阵相等, 即: $\sum_1 = \sum_2 = \dots = \sum_g = \sum$ 。则得到 Bayes 判别函数为

$$W_i(x) = \mu_i^T \sum^{-1} x - \frac{1}{2} \mu_i^T \sum^{-1} \mu_i + \ln q_i \quad (4)$$

实际应用中, 若 $\mu_i, \sum_i$ 未知, 则以训练样本作估计, 即以训练样本的样本均值 $\bar{x}^{(i)}$ 和样本方差 S , 作为 μ_i 和 $\sum_i$ 的估计, 此时,

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^{(k)} = \bar{x}^{(k)} \quad (5)$$
$$\sum = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_g - 1)S_g}{n_1 + n_2 + \dots + n_g - g} \quad (6)$$
$$S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_i^k - \bar{x}^{(k)})(x_i^k - \bar{x}^{(k)})^T \quad (7)$$

式中, $k = 1, 2, \dots, g$ 。

1.4 多正态总体的 Bayes 判别准则

设总体 $G_1, G_2, \dots, G_g$ 的先验概率分布分别为 $q_1, q_2, \dots, q_g$, 误判损失为 $c(j/i)$ ($i, j = 1, 2, \dots, g, i \neq j$)。记 $c(j/i) = 0$ 。在等误判损失下, 其 Bayes 判别准则为

$$x \in G_i, \max_{1 \leq k \leq g} \{ W_i(x) \} = W_i(x) \quad (8)$$

判别准则可解释为: 给定的样品, 计算 g 个总体在 x 处的概率密度函数值 (以 Bayes 函数值 $W_i(x)$ 代替), 最大值所属的总体 G_i 即为 x 所属的类别。

1.5 判别准则的评价

为考察上述判别准则的优良性, 采用以训练样本为基础的刀切法来计算误判率。其基本思想是每次剔除训练样本中的一个样本, 利用其余容量为 $n_1 + n_2 + \dots + n_{g-1}$ 的训练样本建立判别函数, 再用所建立的判别函数对删除的那个样品作判别, 重复上

述步骤,以其误判的比例作为误判概率的估计,则总得误判比例为:

$$\hat{\alpha}_J \triangleq \frac{n_{1M}^J + n_{2M}^J + \cdots + n_{gM}^J}{n_1 + n_2 + \cdots + n_g} \tag{9}$$

式中, n_{gM}^J 表示将属于总体 G_g 的样本误判为其他总体的个数。

2 材料与方法

2.1 野外设计与取样方案

研究区分别设在内蒙古河套灌区的上游(杭锦

后旗),中游(临河)和下游(乌拉特前旗),如图 1。在上游区布设水位观测井 16 眼,中游区布设 14 眼,下游区布设 14 眼,观测井根据地形条件,采用井深 10 m 的普通观测井,井管下端用直径 7.62 cm 塑料管,上部用 1.5 m 长的钢管加帽,并用混凝土加固,防止丢失。

数据采集时间为 2007 年 7 月—2010 年 11 月,每月 21 日取一次水样。

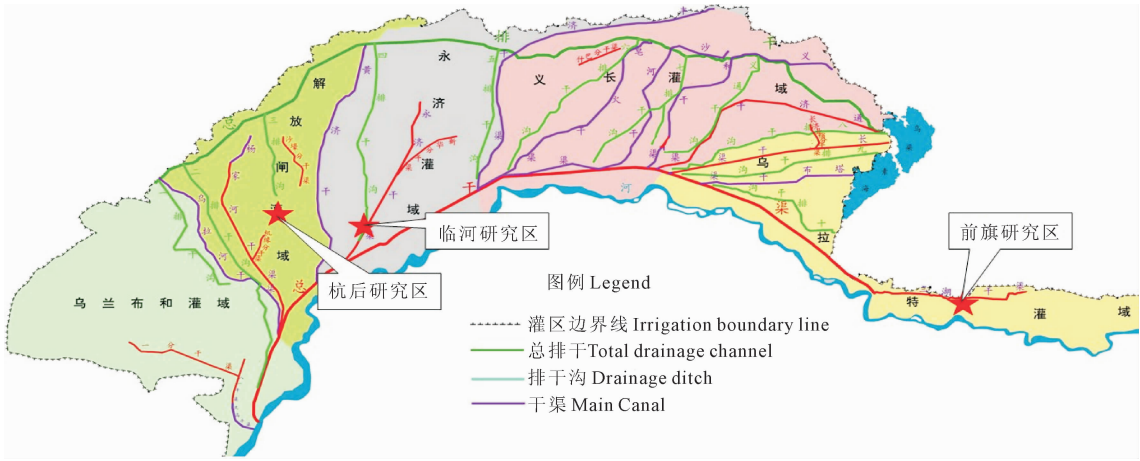


图 1 内蒙古河套灌区研究区分布示意图

Fig.1 Location of Inner Mongolia Hetao Irrigation District in this study

2.2 测试项目及测试方法

对提取水样进行 CO_3^{2-} 、 HCO_3^{-1} 、 Cl^{-} 、 SO_4^{2-} 、 Ca^{2+} 、 Mg^{2+} 、 K^{+} 、 Na^{+} 测定分析,样品测试在内蒙古巴彦淖尔盟水利科学研究所实验中心进行,其中 Na^{+} 、 K^{+} 测试方法为火焰原子吸收分光光度法; Cl^{-} 、 SO_4^{2-} 测试方法为离子色谱法; Ca^{2+} 、 Mg^{2+} 测试方法为 EDTA 滴定法; HCO_3^{-} 、 CO_3^{2-} 测试方法为酸碱滴定法^[13]。

2.3 数据分析

在数据处理过程中采用 SPSS19.0 软件进行 Bayes 判别分析和刀切法评价。

3 结果与分析

3.1 灌区地下水化学类型评判的 Bayes 模型

3.1.1 评价因子 本研究选取 $\text{Na}^{+} + \text{K}^{+}$ (x_1)、 Ca^{2+} (x_2)、 Mg^{2+} (x_3)、 HCO_3^{-} (x_4)、 Cl^{-} (x_5)、 SO_4^{2-}

(x_6)等 6 个指标作为判别因子。原始数据中,地下水水化学组成各离子含量一律取 $\text{mg} \cdot \text{L}^{-1}$ 。为了消除各变量取值大小相差较大时造成对模式识别分类的影响,采用对数变换的方法,即:

$$x_{\text{new}} = \lg(x_{\text{old}}) \tag{10}$$

使变量的变化幅度处在同一水平上。根据多矩形图解法(MRD)的分析结果^[6],将内蒙河套灌区地下水化学类型分为三种类型,即 $\text{Cl}-\text{Na}$ 型(G_1)、 SO_4-Na 型(G_2)和 HCO_3-Na 型(G_3)。将这 3 个类型作为 Bayes 判别分析的 3 个正态总体,以建立水化学类型评判的 Bayes 模型。

3.1.2 Bayes 判别分析函数 参与判别模型建立的地下水样本共计 95 例,其中训练样本集 80 例,预测样本集 15 例。按照第 1 节的方法建立 Bayes 判别函数为:

$$\begin{cases} W_1(x) = -264.737x_1 + 58.911x_2 - 109.891x_3 + 384.178x_4 + 145.030x_5 + 121.896x_6 - 476.720 \\ W_2(x) = -260.684x_1 + 65.926x_2 - 115.987x_3 + 379.444x_4 + 128.019x_5 + 138.901x_6 - 478.802 \\ W_3(x) = -256.641x_1 + 67.327x_2 - 107.281x_3 + 393.264x_4 + 127.359x_5 + 116.864x_6 - 485.418 \end{cases} \tag{11}$$

以判别准则对上述判别函数的优良性进行检验,总误判率为 17.5%,识别正确率达 82.5%。

3.2 Bayes 判别模型的应用

应用训练后的模型对河套灌区 15 个地下水水样进行水化学类型的分类识别。将表 1 中的预测样本数据分别代入式(11)中计算 $W_i(x)(i=1,2,3)$,

计算结果见表 2。按式(8)进行判别,以 $W_i(x)$ 中的最大值对应的总体为样品所归属的总体,评判结果见表 2,从表中可知判别模型对 15 个地下水预测样本进行了识别,其中有 13 个识别正确,2 个识别失误,正确率达 86.67%,所建立模型正确率较高可应用于实际当中。

表 1 河套灌区地下水化学类型分类指标实测值
Table 1 Measurement values of hydrochemical types in underground water of Hetao Irrigation District

预测样本 Forecast samples	评价指标实测值 Evaluation index						分类结果 Classification
	HCO ₃ ⁻	CL ⁻	SO ₄ ²⁻	Ca ²⁺	Mg ²⁺	Na ⁺ + K ⁺	
1	2.886	2.720	2.764	1.997	2.087	2.755	Cl - Na 型
2	2.660	2.565	2.603	1.930	1.954	2.533	Cl - Na 型
3	2.750	2.658	2.336	2.117	1.691	2.487	Cl - Na 型
4	2.621	2.594	2.719	1.836	1.997	2.672	Cl - Na 型
5	2.754	2.578	2.555	2.008	1.713	2.612	Cl - Na 型
6	2.934	3.194	3.072	1.915	2.182	3.184	Cl - Na 型
7	2.640	2.481	2.911	2.184	2.141	2.497	SO ₄ - Na 型
8	2.784	2.302	2.778	2.091	1.933	2.535	SO ₄ - Na 型
9	2.661	2.124	2.301	1.93	1.811	2.164	HCO - Na 型
10	2.531	2.078	2.017	1.662	1.562	2.126	HCO - Na 型
11	2.769	2.270	2.453	1.979	1.844	2.407	HCO - Na 型
12	2.615	2.334	2.358	1.939	1.549	2.412	HCO - Na 型
13	2.866	2.597	2.527	2.084	2.162	2.454	HCO - Na 型
14	2.725	2.345	2.516	1.904	2.035	2.346	HCO - Na 型
15	2.551	2.290	2.301	2.062	1.730	2.125	HCO - Na 型

表 2 地下水样品的 Bayes 判别结果
Table 2 Bayes discriminant results on underground water samples

预测样本 Forecast samples	判别函数值 Discriminant scores				归属 Affiliation	水化学类型 Hydrochemical type
	$W_1(x)$	$W_2(x)$	$W_3(x)$	$Max(W(x))$		
1	522.239	519.683	522.338	522.338	G ₃	HCO - Na 型
2	462.810	460.671	461.708	462.810	G ₁	Cl - Na 型
3	530.421	524.453	530.357	530.421	G ₁	Cl - Na 型
4	419.434	418.588	417.317	419.434	G ₁	Cl - Na 型
5	505.017	503.739	505.447	505.447	G ₃	HCO - Na 型
6	518.105	513.091	511.754	518.105	G ₁	Cl - Na 型
7	484.672	489.773	485.651	489.773	G ₂	SO ₄ - Na 型
8	504.945	510.932	510.049	510.932	G ₂	SO ₄ - Na 型
9	475.694	475.289	480.555	480.555	G ₃	HCO - Na 型
10	406.442	402.086	409.150	409.150	G ₃	HCO - Na 型
11	491.960	492.270	496.916	496.916	G ₃	HCO - Na 型
12	459.073	458.942	460.918	460.918	G ₃	HCO - Na 型
13	544.414	538.934	546.201	546.201	G ₃	HCO - Na 型
14	484.651	483.016	486.936	486.936	G ₃	HCO - Na 型
15	484.934	483.483	486.440	486.440	G ₃	HCO - Na 型

4 小结与讨论

本文应用 Bayes 判别分析理论,选用 $\text{Na}^+ + \text{K}^+$ 、 Ca^{2+} 、 Mg^{2+} 、 HCO_3^- 、 Cl^- 、 SO_4^{2-} 等 6 个指标作为判别因子,建立了地下水化学类型的 Bayes 判别分类模型。所建立的判别模型采用刀切法对 Bayes 判别准则的优良性进行评价,识别正确率达 82.5%。同时,以该模型对预测样本进行判别分析,输出结果正确率达 86.6%,从而验证了 Bayes 判别准则的可行性和可靠性。

与传统的分类方法相比,Bayes 判别分类法存在误判的情况,它是一种带有预测性的判别分类方法,其准确度依赖于样本的数量,在建模过程中会受样本数据代表性和准确性的影响。因此在使用判别分类法时,对样本数量和样本的代表性有一定的要求。

本研究判别模型的建立是以多矩形图解(MRD)法为前提,所以与传统分类方法相比,该模型提供的判别分类结果,具有更明晰、易懂的水化学类型信息。

该模型建立所依据的数据来自内蒙古河套灌区,所以模型适用于和河套灌区相似的北方盐渍化灌区。模型简单易操作,准确率较高,满足工程领域要求,在今后的使用过程中,随着训练样本容量的增大,其代表性和准确性不断提高,模型逐步完善,可见 Bayes 判别分类法具有较高的应用价值,为地下水化学类型评判与分类提供了新的数量化和定量化途径。

参 考 文 献:

[1] 杨天笑.对现有地下水化学分类的几点认识[J].华东地质学院学报,1995,18(1):40-42.

[2] 李育松,胡显欣,卞建民.模糊聚类分析在地下水化学类型分类中的应用[J].人民黄河,2013,35(7):37-39,43.

[3] 朱学愚,钱孝星.地下水水文学[M].北京:中国环境科学出版社,2005.

[4] 寇文杰.基于 EXCEL 的地下水化学舒卡列夫分类方法[J].工程勘察,2013,(5):48-50.

[5] 孙亚乔,钱 会,张 黎,等.银川地区地下水化学特征[J].干旱地区农业研究,2006,24(3):185-189.

[6] 周金龙.多矩形图解法及其在塔里木盆地中的应用[J].水文地质工程地质,2006,(2):4-6.

[7] 朱海勇,陈永金,刘加珍,等.塔里木河中下游地下水化学及其演变特征分析[J].干旱区地理,2013,36(1):8-18.

[8] 张庆利.SPSS 宝典[M].第二版.北京:电子工业出版社,2011.

[9] 梁 才.基于 Bayes 判别理论的钢筋混凝土施工质量综合评价方法[J].混凝土,2011,(5):50-52.

[10] 梁仕华,文畅平.膨胀土判别与分类的 Bayes 判别分析法[J].力学与实践,2008,30(6):73-76.

[11] 文畅平.岩体质量分级的 Bayes 判别分析方法[J].煤炭学报,2008,33(4):395-399.

[12] 宫凤强,李夕兵,张 伟.基于 Bayes 判别分析方法的地下工程岩爆发生及烈度分级预测[J].岩土力学,2010,31(S1):370-377.

[13] 李 彬,史海滨,闫建文,等.节水改造后盐渍化灌区区域地下水埋深与土壤水盐的关系[J].水土保持学报,2014,28(1):117-122.

[14] 李韵珠,陆锦文,吴金绥.河北曲周盐渍土区的地下水化学特征[J].土壤学报,2001,38(3):241-352.

[15] 马 雷,钱家忠,赵卫东.基于 GIS 的地下水化学类型空间分区方法[J].煤炭学报,2012,37(3):490-494.

[16] 虎胆·吐马尔白,杨路华,史海滨,等.地下水利用[M].北京:中国水利水电出版社,2008.

(上接第 104 页)

[17] Ierna A, Lombardo G P, Mauromicale G. Tuber yield, water and fertilizer productivity in early potato as affected by a combination of irrigation and fertilization[J]. Agricultural Water Management, 2011, 101(1):35-41.

[18] 郑顺林,李首成,丁海萍.高强力优质棉干物质积累和分配规律的研究[J].湖北农业科学,2006,45(1):44-48.

[19] 伍维模,郑德明,董合林.新疆棉花干物质和氮磷钾养分积累的模拟分析[J].西北农业学报,2002,11(1):92-96.

[20] 张旺锋,李蒙春,勾 玲,等.北疆高产棉花养分吸收特性的研究[J].棉花学报,1998,10(2):88-95.

[21] 薛晓萍,郭文琦,王以琳,等.不同施氮水平下棉花生物量动态增长特征研究[J].棉花学报,2006,18(6):323-326.

[22] FB Fritschi, BA Roberts, RL Travis, et al. Response of irrigated a-cala and pima cotton to nitrogen fertilization[J]. Agronomy Journal, 2003,95:133-146.

[23] 李培岭,张富仓.根系分区交替滴灌下水氮耦合对棉花氮素利用效率的影响[J].农业工程学报,2012,28(25):112-116.

[24] 薛晓萍,王建国,郭文琦,等.氮素水平对初花后棉株生物量、氮素累积特征及氮素利用率动态变化的影响[J].生态学报,2006,26(11):3631-3640.

[25] 谢志良,田长彦.膜下滴灌水氮耦合对棉花干物质积累和氮素吸收及水氮利用效率的影响[J].植物营养与肥料学报,2011,17(1):160-165.

[26] 邓 忠,白 丹,翟国亮,等.膜下滴灌水氮调控对新疆棉花产量及水氮利用率的影响[J].应用生态学报,2013,24(9):2525-2532.

[27] 周立平,郭金强.水氮耦合效应对膜下滴灌棉花产量的影响[J].沙漠与绿洲气象,2012,6(6):51-54.