

农作区净灌溉需水量模拟及不确定性分析

马黎华¹, 康绍忠^{2*}, 粟晓玲¹, 佟玲²

(1. 西北农林科技大学水利与建筑工程学院, 杨凌 712100; 2. 中国农业大学中国农业水问题研究中心, 北京 100083)

摘要: 净灌溉需水量是估算农业灌溉用水量的参考依据。该文以西北干旱内陆区石羊河流域上中下游的古浪县, 凉州区, 民勤县为研究对象, 在分析农业净灌溉需水量宏观驱动力因子(1959—2005)的基础上, 以关键因子作为输入项, 区域农业净灌溉量为输出项, 分别建立农业净灌溉需水量的多元线性回归模型、人工神经网络 BP 模型以及人工神经网络集成模型。并对不同模型的模拟效果进行比较; 通过对时间序列的离散化蒙特卡洛(MC)设计, 采用不确定性评价指数(d-factor)对 3 种模型模拟的不确定性进行分析。结果表明: 与多元线性回归模型和神经网络 BP 模型相比, 神经网络集成模型具有较高的模拟精度, 并能合理地指示影响因素与净灌溉需水量的不确定性变化。

关键词: 灌溉, 不确定性分析, 蒙特卡洛模拟, 多元线性回归模型, 人工神经网络集成模型

doi: 10.3969/j.issn.1002-6819.2012.08.002

中图分类号: S152.7

文献标志码: A

文章编号: 1002-6819(2012)-08-0011-08

马黎华, 康绍忠, 粟晓玲, 等. 农作区净灌溉需水量模拟及不确定性分析[J]. 农业工程学报, 2012, 28(8): 11—18.

Ma Lihua, Kang Shaozhong, Su Xiaoling, et al. Simulation and uncertainty analysis of net irrigation requirement in agricultural area[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(8): 11—18. (in Chinese with English abstract)

0 引言

净灌溉需水量是指必须通过灌溉补充的土壤原有储水量和有效降雨量及地下水利用量不能满足作物蒸发蒸腾、冲洗盐碱以及其他方面要求的水量^[1]。净灌溉需水量是农业灌溉用水量的参考依据^[2-3], 可以通过计算来实现。一般按照一定的物理规律和田间水量平衡原则计算以天为步长的区域灌溉需水量^[4]; 农业净灌溉需水量的计算所需的灌溉面积, 作物耕种面积, 作物系数, 气象因素等指标量往往难以统计, 存在不确定性因素, 因此实现农业净灌溉需水量的精确计算往往比较困难。外部环境的变化以及政策的调整会对农业净灌溉需水量产生不同程度的影响, 从影响因素入手实现对农业净灌溉需水量的模拟对于了解区域农业用水状况以及实现区域农业用水的预测具有积极的现实意义。在以往农业净灌溉需水量的影响因素分析中, 所涉及到的影响关系复杂、影响因素繁多; 其中人口^[5]、全球气候变化、种植结构的调整^[6]、以及水价与农药消耗等经济因素^[7]都被认为会对净灌溉需水量造成一定的影响。以关键的影响因素为输入, 以农业净灌溉需水量为输出建立输入输出的农业净灌溉

需水量模型的研究还较少见。以数字驱动的输入输出模型有多种形式, 其中线性模型简单且易于实现, 并由于系数对变量的解释性被广泛应用。在未知输入与输出项之间的关系是线性还是非线性的时候, 人工神经网络模型常作为线性模型的对比与补充^[8-9]。在数据样本数量较少的情况下(低于 200 个), ANN 模型在训练过程中易出现收敛不稳定, 神经网络集成旨在通过训练多个神经网络并将其进行组合来提高神经网络系统地泛化能力^[10-11]。交叉验证(Cross Validation, CV)常常用于考察模型的模拟能力。基于数字驱动的输入输出模型注重变量之间的先验判断和统计关系, 往往只能实现对数据样本相对准确的模拟, 所得到的模型在应用时存在一定的不确定性, 在一定范围内的输入项的变化对于模型结果的影响成为衡量模型这种不确定性的标准。针对这种不确定性的分析, 有助于实现对于模型的合理选择和准确应用。不确定性分析一般通过蒙特卡洛(Monte-Carlo, MC)模拟来实现, MC 模拟在水文以及水资源管理等领域的确定性分析中有着广泛的应用, 其中以针对参数的不确定性分析与针对输入项不确定性分析最为常见^[12-14]。

在干旱内陆区石羊河流域, 缺水不仅带来了严重的生态环境问题, 更是限制了区域社会经济的发展^[3]。在这种背景下, 石羊河流域农业用水占总用水量的比例却高达 90%(1998—2008 的平均值), 本文根据上中下游不同的农业净灌溉需水量的特性, 以石羊河流域上中下游的古浪县, 凉州区, 民勤县 3 个研究区域 45 a 的数据为研究对象; 并对潜在的影响因素(1959—2005)进行分析, 选择关键因子; 将选择的关键因子作为输入项, 农业净灌溉需水量为输出项, 建立农业净灌溉需水量的多元线性回归模型, 人工神经网络模型与人工神经网络集

收稿日期: 2012-03-12 修订日期: 2012-03-31

基金项目: 国家高技术研究发展计划(863 计划)(2011AA100502); 国家自然科学基金青年基金项目(51109211); 西北农林科技大学基本科研业务费科技创新重点项目(编号 QN201168)

作者简介: 马黎华(1983—), 女, 河南平顶山人, 博士生, 主要从事农业节水与水资源领域的研究。杨凌 西北农林科技大学水利与建筑工程学院, 712100。Email: malihua1210@yeah.net

※通信作者: 康绍忠(1962—), 男, 湖南桃源人, 教授, 博士生导师, 主要从事农业节水与水资源领域的研究。北京 中国农业大学中国农业水问题研究中心, 100083。Email: kangshaozhong@tom.com

成模型; 基于年为步长时间序列的离散化蒙特卡洛 (MC) 设计对模型进行不确定分析, 通过模型的误差结果与不确定性分析比较不同模型在应用于区域农业净灌溉需水量模拟的优劣。

1 材料与方法

1.1 研究区域

石羊河流域是甘肃省河西走廊三大内陆河流域之一, 位于祁连山东段与巴丹吉林沙漠、腾格里沙漠南缘之间, E 101°41'~104°16'、N 36°29'~39°27', 流域面积 4.16 万 km²。流域水资源匮乏, 常年干旱少雨、日照强烈、蒸发量大, 具有大陆性温带干旱气候特征。流域水资源量为 16.608 亿 m³, 2008 年总用水量为 26.86 亿 m³, 其中农业灌溉用水高达 23.21 亿 m³, 占总用水量的 86.43%。目前流域共有 8.6 万 hm² 农田面积得不到灌溉, 灌溉用水与其他部门用水之间矛盾十分尖锐。

石羊河流域地势南高北低, 由西南到东北方向依次为上游祁连山区, 中游武威盆地, 下游民勤盆地。上游祁连山区以山地为主, 兼有耕地种植。近年来, 流域上游祁连山区由于人为砍伐森林, 过度放牧, 开矿挖药和开荒种植, 植被破坏严重, 山区的植被覆盖率只有 40% 左右。中游武威盆地有悠久的耕种历史, 其灌溉面积持续扩大。据统计, 中游灌溉面积由解放初期的 165 万 hm² 增加到 2003 年的 278 万 hm², 增长了 1.7 倍; 近 50 a 来, 总耗水量由解放初的 5.67 亿 m³ 增加到 2003 年的约 10 亿 m³。下游民勤盆地的社会经济也是以农耕为主, 其地表水源红崖山水库的供给主要来自于中游的退水、余水和上游的洪水。近 50 a 来红崖山入库断面平均年径流量 (主要为下游民勤盆地提供灌溉用水) 由 4.6 亿 m³, 锐减为现状不足 1.0 亿 m³[15]。由以上分析可知, 上中下游的农业用水呈现不同的特性, 针对上中下游分别建立农业用水模型研究区耕地分布与行政区划如图 1 所示。

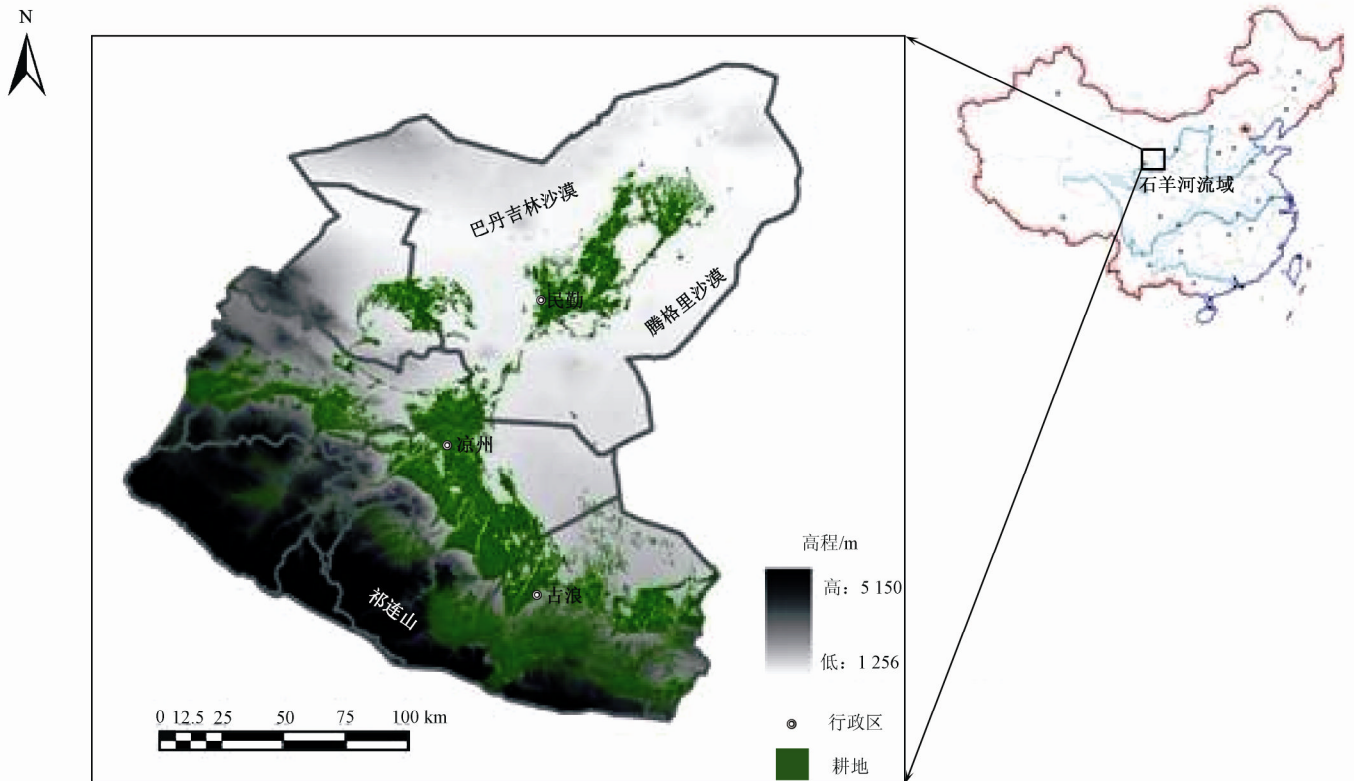


图 1 石羊河流域高程与耕地分布图

Fig.1 Distribution of cultivated land in Shiyang river basin and its digital elevation model (DEM)

1.2 数据准备

1.2.1 农业净灌溉需水量

本文采用的农业净灌溉需水量数据来源于佟玲[16]关于石羊河流域多年净灌溉需水量的计算结果。该值的计算原理如下

$$I_N = \frac{1}{1000} \sum (I_i \times A_i) \quad (1)$$

式中, I_N 为农业总净灌溉需水量 (10⁸m³); I_i 为第 i 种作物全生育期净灌溉需水量, mm; A_i 是作物的种植面积, 万 hm²。

其中各作物净灌溉需水量 I_i 采用下式计算

$$I_i = ET_c - P_e \quad (2)$$

式中, ET_c 为作物蒸发蒸腾量, mm; P_e 为作物生育期内有效降雨量, mm。

各作物 ET_c 采用作物系数与参考作物蒸发蒸腾量相乘

$$ET_c = K_c \times ET_0 \quad (3)$$

式中, K_c 是作物系数, 由石羊河流域相关试验数据得到。 ET_0 为作物蒸发蒸腾量, mm; 根据气象资料, 采用 1998 年 FAO 灌溉排水丛书第 56 分册推荐的 Penman-Monteith 公式计算各站历年各月 ET_0 [17]。

1959—2003 年的古浪, 凉州与民勤的农业净灌溉需水量变化如图 2 所示。对比 3 个地区, 凉州的农业净灌溉需水量

最大，尽管 1970 年前呈下降趋势，但 1970 年后增长迅速。古浪与民勤农业净灌溉需水量变化不大，近年呈下降趋势。

1.2.2 潜在影响因素

潜在的影响因素既要与农业净灌溉需水量密切相关，同时又要考虑数据来源的可靠性以及完备性。选择的初始指标有：人口（ 10^4 ），城镇化率（城镇人口比例），有效灌溉面积（ 10^4 hm^2 ），耕地面积（ 10^4 hm^2 ），粮食总产量（ 10^4 t ），农业产值（ 10^8 元 ），工业产值（ 10^8 元 ），国内生产总值（GDP）（ 10^8 元 ），经济作物比例，降水（mm），气温（ $^{\circ}\text{C}$ ）， ET_0 （mm）共计 12 项。其中气象指标数据来自区域内国家气象站点的统计（1959—2005）； ET_0 数据参照佟玲（2007）^[17]的相关文献；社会经济指标数据来自甘肃省相关的统计年鉴（1959—2005）。指标间的相关系数如表 1 所示。

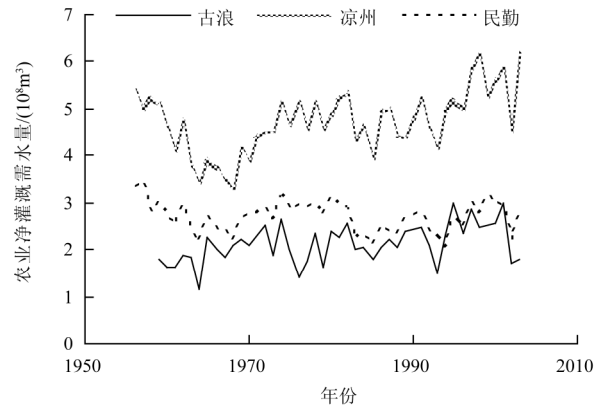


图 2 古浪、凉州与民勤农业净灌溉需水量的多年变化图（1959—2003）

Fig.2 Changes of net irrigation requirements in Gulang, Liangzhou and Minqin during 1959—2003

表 1 影响因素之间的相关分析

Table 1 Results of the correlation analysis from the potential factors

指标	人口	城镇化率	灌溉面积	耕地面积	粮食产量	农业产值	工业产值	GDP	经济作物比例	降水	温度	ET_0
人口	1											
城镇化率	0.908**	1										
灌溉面积	0.438**	0.608**	1									
耕地面积	0.593**	0.702**	0.649**	1								
粮食产量	0.968*	0.869**	0.418**	0.574**	1							
农业产值	0.814*	0.898**	0.655**	0.804**	0.820**	1						
工业产值	0.866**	0.912**	0.564**	0.780**	0.856**	0.977**	1					
GDP	0.810**	0.911**	0.656**	0.746**	0.807**	0.983**	0.962**	1				
经济作物比例	0.936**	0.939**	0.496**	0.647**	0.920**	0.865**	0.891**	0.859**	1			
降水	0.271	0.190	0.049	0.079	0.291	0.227	0.227	0.244	0.224	1		
温度	0.616**	0.651**	0.541**	0.653**	0.569**	0.701**	0.721**	0.698**	0.657**	0.013	1	
ET_0	0.032	0.056	0.209	0.279	0.002	0.143	0.124	0.084	0.092	-0.577**	0.436**	1

注：*表示达到 0.05 显著性水平， $R_{0.05}=0.304$ ；**表示达到 0.01 的显著性水平 $R_{0.01}=0.4$ 。

1.3 模型建立与不确定性分析方法

1.3.1 输入项的选择

模型输入项的选择遵循独立原则，尽量避免输入项之间存在重要的依存关系（相关关系）；同时互信息（Mutual Information, MI）能够作为输入项选择的一种有效指标^[18-20]。按照相关系数与互信息的计算原理，分别计算农业净灌溉需水量与潜在影响因素之间的相关系数与互信息值。为了消除不同影响因子之间的量纲差异，影响因素与净灌溉需水量都经过了正态标准化处理。

基于线性关系的相关分析是通过计算相关系数并根据自由度以及显著性标准的比较来判断两个变量之间的相关关系（R）。通过与 R 在显著性水平值的对比，例如 $R \geq R_{0.05}$ 为显著性相关， $R \geq R_{0.01}$ 为特显著性相关；R 值越大该因素与用水结构之间的关系越紧密。

从而影响因素的时间序列为 $H(X)$ ，净灌溉需水量的时间序列为 $H(Y)$ ，二者在不同的变化过程中一致性变化部分可以由互信息（MI）表示，MI 函数用于描述两个随机变量 X 与 Y 之间的关系

$$MI = \iint f_{X,Y} \ln \left[\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)} \right] dx dy \quad (4)$$

其中， $f_X(x)$ ， $f_Y(y)$ 是变量 X, Y 的边际分布函数（probability distribution functions）。 $f_{X,Y}(x,y)$ 是 X, Y 的联合（二元）边际分布函数。如果 X, Y 之间无关系，则 $f_{X,Y}(x,y)$ 的值与 $f_X(x)$ ， $f_Y(y)$ 的值相等，MI 的值为 0；反之，如果随机变量 X, Y 关系较紧密，则 MI 的值较大。

潜在影响因素与石羊河流域农业净灌溉需水量的相关系数与互信息（MI）值的计算结果如表 2 所示。由表 2 可以看出，除降水指标外，所选择的影响因素与区域农业净灌溉需水量之间均存在显著的相关关系。其中 ET_0 为显著相关；人口，城市化率，粮食产量，GDP，农业产值，工业产值，有效灌溉面积，耕地面积以及经济作物比例为特显著相关。由 MI 计算结果可以看出除了城镇化率与经济作物比例，其他指标与农业净灌溉需水量之间的 MI 值较大。

在影响因素的选择中，以 MI 的结果为基础，参照指标间的相关关系结果（表 1）进行选择。例如，人口与城市化率都是与区域的人口密切相关的量，当二者之间存在特显著相关关系（0.908**）（表 1），参照 MI 的结果，只选择人口（具有较大的 MI 值的）作为此分类中的选择结果。灌溉面积与耕地面积都属于面积因素，二者之间有特显著相关关系，耕地面积的 MI 值大于灌溉面积的

MI 值，因此选择耕地面积作为此分类中的选择结果。粮食产量与农业产值，工业产值，GDP，经济作物比例之间都有特显著相关的关系，这些指标均属于与经济密切相关的量，根据粮食产量在其中具有最高的 MI 值，选取粮食产量

作为这一分类的代表。同理， ET_0 与降水，气温存在特相关关系，在气象指标分类中具有最高的 MI 值，选取 ET_0 。因此，最终选取的关键因子，即数字驱动力模型的输入项有人口，耕地面积，粮食产量和 ET_0 ，共计 4 个指标量。

表 2 农业净灌溉需水量与影响因素的相关分析(R)与互信息值(MI)结果列表
Table 2 Results of the correlation analysis and mutual information (MI) method

	人口	城镇化率	灌溉面积	耕地面积	粮食产量	农业产值	工业产值	GDP	经济作物比例	降水	温度	ET_0
净灌溉需水量 (R)	0.909**	0.869**	0.513**	0.699**	0.898**	0.869**	0.875**	0.837**	0.905**	0.099	0.660**	0.311*
净灌溉需水量 (MI)	2.654	0	0.557	0.777	2.637	1.689	2.113	2.174	0	2.371	0.619	2.443

注：*表示达到 0.05 显著性水平， $R_{0.05}=0.304$ ；**表示达到 0.01 的显著性水平 $R_{0.01}=0.4$ 。

1.3.2 模型的建立方法

由输入项的选择结果，基于数据驱动的输入输出模型在模拟石羊河流域农业净灌溉需水量变化中采用了三种不同类型的模型，分别是①多元线性回归模型 (Multiple-Linear Regression, MLR)；②人工神经网络模型 (BP 算法) (Artificial Neural Network, ANN)；③人工神经网络集成模型 (Ensemble ANN)。数据采用原始数值和量纲，所有模型的输出项均为净灌溉需水量的模拟值 ($10^8 m^3$)，输入项均为人口 (10^4)，粮食产量 ($10^4 t$)，耕地面积 ($10^4 hm^2$)，气象指标 (mm)。

神经网络模型应用广泛，在实际应用中往往需要经过大量的实验摸索才能够确定合适的神经网络模型、算法以及参数设置。神经网络集成模型 (Ensemble ANN) 是“用有限个神经网络对同一个问题进行学习，集成在某输入示例下的输出由构成集成的各神经网络在该示例下的输出共同决定”^[11]，其易于使用且效果明显，被视为工程化的神经计算方法。本文中神经网络集成模型采用 Bagging 算法^[21]，个体神经网络是在同样数据样本、同样的 BP (Back Propagation) 算法下随机得到的不同网络。

神经网络集成模型原理如下，网络结构示意图如图 3 所示。

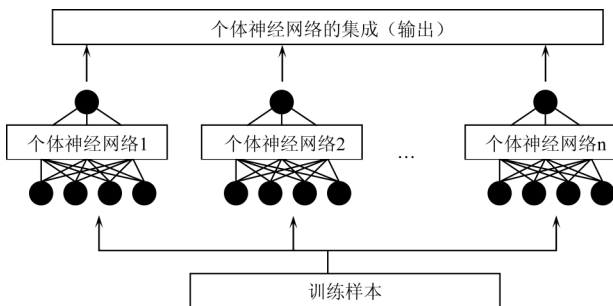


图 3 神经网络集成模型的建模流程示意图

Fig.3 Framework of ensemble artificial neural network model

$$\hat{I}_{-nn} = \frac{1}{n} \sum_{a=1}^n ANN_a(x_1, x_2, x_3, x_4) \quad (5)$$

其中， \hat{I}_{-nn} 为区域农业净灌溉需水量的模拟值，由选择后的输入项决定， $10^8 m^3$ ； a 为个体神经网络的个数； x_1 为人口， 10^4 ； x_2 为粮食产量， $10^4 t$ ； x_3 为耕地面积， $10^4 hm^2$ ； x_4 为 ET_0 ，mm。

1.3.3 模型的选择与评价

数字驱动模型的假设之一，就是数据样本中包含了有价值的信息（变化规律）；可以认为随时间的变化，输入与输出之间的关系发生变化，而这种变化趋势被模型记录并实现模拟。对于输入输出模型为数字驱动模型，由于样本数量较少，训练样本的选择会对模型的结果造成较大的影响^[22]。因此在最优模型的选择与比较中多采用交叉验证 (Cross Validation, CV) 的方式，尤其对于神经网络等类似黑箱模型更为适用^[23-25]。

模型拟合误差的计算采用对检测样本的均方根误差 (Root Mean Square Error, RMSE) 和预测残差平方和 (Predicted Residual Error Sum of Squares, PRESS)，其中 PRESS 多用于交叉验证，RMSE 多用于模型模拟的效果的比较。其计算公式如下

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (I_j - \hat{I}_j)^2} \quad (6)$$

$$PRESS = \sum_{j=1}^n (I_j - \hat{I}_j)^2 \quad (7)$$

其中， I_j 是农业净灌溉需水量 ($10^8 m^3$)， \hat{I}_j 是由模型模拟的农业净灌溉需水量 ($10^8 m^3$)， n 为样本数量， j 为第 j 年个样本。

1.3.4 不确定性分析的方法

对已建立模型的不确定性分析，采用时间序列离散化的蒙特卡洛 (MC) 设计。离散化本来指将连续问题的解用一组离散要素来表征而近似求解的方法，而在实际应用中离散化其实是程序设计中一个常用的技巧，它可以有效的降低时间复杂度。本文所用的离散化 MC 设计就是将 2 a 的数据作为阈值，根据平均分布的分布函数随机得到离散化的数据样本 (输入项) 做为新的输入项 (时间序列)，再将这些输入项导入已经建立的区域农业净灌溉需水量模型中，对模型的输出结果再进行进一步的分析。

标准的平均分布 U 的概率密度函数 (probability density function, PDF) 为

$$f(u) = \begin{cases} 1 \\ 0 \end{cases} \quad (8)$$

由于密度是连续的， U 在 $[0, 1]$ 的范围内均匀分布。在此基础上，当 $W=(b-a)U+a$ ，可以推出在 $[a, b]$ 范围内

的平均分布 W ，则 W 的概率密度函数 (PDF) 为

$$g(w) = \begin{cases} \frac{1}{b-a} \\ 0 \end{cases} \quad (9)$$

其中，当 W_i 为第 i 年的输入项时有： $a=w_{i-1}$ ， $b=w_{i+1}$ 。

本文中采用 d-factor 指数^[26-27]对 MLR 模型，ANN 模型以及 Ensemble ANN 模型的不确定性进行比较。其公式为

$$\bar{d}_I = \frac{1}{n} \sum_{j=1}^n (I_{Uj} - I_{Lj}) \quad (10)$$

$$d\text{-factor} = \frac{\bar{d}_I}{\sigma_I} \quad (11)$$

其中， \bar{d}_I 为指示性的上限 (I_U) 与下限 (I_L) 的平均距离，即置信区间之间的平均距离； n 为数据组的数量； σ_I 为观测值的标准差。其中，上限 (I_U) 与下限 (I_L) 的值由 MC 模拟得到的时间序列中每一组数据的 95%机会数与 5%机会数决定。d-factor 值越大则模型进行 MC 模拟的不确定性越大。

2 结果与分析

2.1 模型的结果与误差分析

2.1.1 神经网络集成模型

k 折交叉验证 (k -fold CV) 是指将所有的样本随机分割为等大的 k 个子集， $k-1$ 的集合作为训练样本， k 作为检测样本；LOO (leave one out) 是其中的极端形式，LOO 的优点是每一个样本都有机会单独作为检测样本^[28]。在不同 k 值下的神经网络集成模型检测样本的预测残差平方和 (PRESS) 如表 3 所示。由表 3 可以看出随着训练样本与检测样本的不同模型模拟效果的波动变化，即随着 k 值得增加，PRESS 先减少再增加然后又减少。比较所需个体神经网络的个数 (表 4)，认为在 $k=5$ 时，所需的子集较少，检测样本误差也较小。

2.1.2 多元线性回归模型

多元线性回归模型结果如表 5 所示。其中，4 个自变量对于因变量的解释充足度以 F 检验表示，单个自变量系数的显著性由 t 检验来判断。结果表明：3 个地区均满足 F 检验的特显著标准，由用人口，耕地面积，粮食产量以及 ET_0 可以实现对 3 个区域的农业净灌溉需水量的模拟；对于不同自变量系数进行 t 检验结果表明，耕地面积与 ET_0 的系数均达到了显著的标准，在民勤地区，自变量中的人口项也达到了显著标准。

表 3 不同 k 取值下古浪，凉州与民勤的农业净灌溉需水量的神经网络集成模型检测样本的预测残差平方和(PRESS)

Table 3 Predicted residual sum of squares (PRESS) of testing samples under Ensemble ANN model for net irrigation requirements in Gulang, liangzhou and Minqin in Shiyang river basin

地区	$k=3$	$k=5$	$k=9$	LOO CV
古浪	4.600	3.810	3.928	2.891
凉州	10.954	8.206	9.582	5.808
民勤	4.900	3.123	3.086	1.964

注： $k=3$ ， $k=5$ ， $k=9$ 与 LOO CV 为交叉验证(k -fold CV) 的 4 种形式。 k 表示样本随机分割为等大的 k 个子集。下同。

表 4 不同 k 取值下农业净灌溉需水量的神经网络集成模型的预测残差平方和(PRESS)与个体神经网络数量的比较

Table 4 Average predicted residual sum of squares (PRESS) of testing samples and the total nets number under Ensemble ANN model for net irrigation requirements in Gulang, liangzhou and Minqin in Shiyang river basin¹

k 值的选择	PRESS	个体神经网络的个数
$k=3$	6.818	15
$k=5$	5.046	25
$k=9$	5.532	45
LOO CV	3.554	225

表 5 古浪，凉州，民勤的多元线性回归模型的结果
Table 5 MLR models for net irrigation requirements in Gulang, liangzhou and Minqin in Shiyang river basin.

地区	F	自变量	系数
古浪	13.442**	人口	-0.0163
		耕地面积	1.5062**
		粮食产量	-0.0001
		ET_0	0.0011**
凉州	27.413**	人口	-0.0283
		耕地面积	0.7940*
		粮食产量	0.0003
		ET_0	0.0015**
民勤	5.058**	人口	0.1741**
		耕地面积	-0.6345*
		粮食产量	-0.0002
		ET_0	0.0006**

注：*表示达到 0.05 显著性水平： $F_{0.95}(4, 46)=2.57$ ， $t_{0.95}(45)=1.6794$ ；**表示达到 0.01 的显著性水平 $F_{0.95}(4, 46)=3.76$ ； $t_{0.99}(45)=2.4121$ 。

2.1.3 多个模型的误差比较

均方根误差越小，模拟值与真实值的差的平方和越小，模型的模拟效果越好。表 6 为检测样本相同时的 3 种模型的误差比较。从 3 种模型的对比来看神经网络集成模型 (Ensemble ANN) 优于神经网络 (ANN) 模型与多元线性回归模型 (MLR)。从地区的对比来看，与上游古浪区与下游民勤县相比，中游凉州区的模拟效果较差。中游凉州地区近 50a 来农业面积增加显著，净灌溉需水量与输入项之间的关系存在较大变化；相比较而言，上游山区的古浪与下游沙漠化情况严重的民勤农业面积与农业规模的变化都较小。可以认为随时间的变化，中游凉州地区的输入与输出之间的关系变化较大，上游古浪与下游民勤的输入与输出之间的关系变化不大。

表 6 古浪，凉州，民勤的多个模型检测样本的误差(RMSE)
Table 6 RMSE of testing samples under MLR, ANN, and Ensemble ANN model for net irrigation requirements in Gulang, liangzhou and Minqin

地区	均方根误差 RMSE		
	MLR	ANN	Ensemble ANN
古浪	0.328	0.335	0.208
凉州	0.454	0.546	0.267
民勤	0.401	0.397	0.143

注：MLR 为多元线性回归模型，ANN 为人工神经网络模型，Ensemble ANN 模型为神经网络集成模型。下同。

2.2 不确定性分析的结果

从不确定性指数的计算结果看(表7):下游民勤地区的 d-factor 值最大,对于输入项的变化最为敏感;从三种模型的对比来看(平均值),神经网络(ANN)模型的 d-factor 值最大,神经网络集成(Ensemble ANN)模型的 d-factor 值最小。

表7 不同模型的不确定性指数(d-factor)

Table 7 Uncertainty measuring parameter (d-factor) for different models in Gulang, Liangzhou, Minqin and Shiyang river basin

	古浪	凉州	民勤	平均值
MLR	0.288	0.239	0.413	0.313
ANN	0.458	0.412	0.621	0.497
Ensemble ANN	0.235	0.217	0.372	0.275

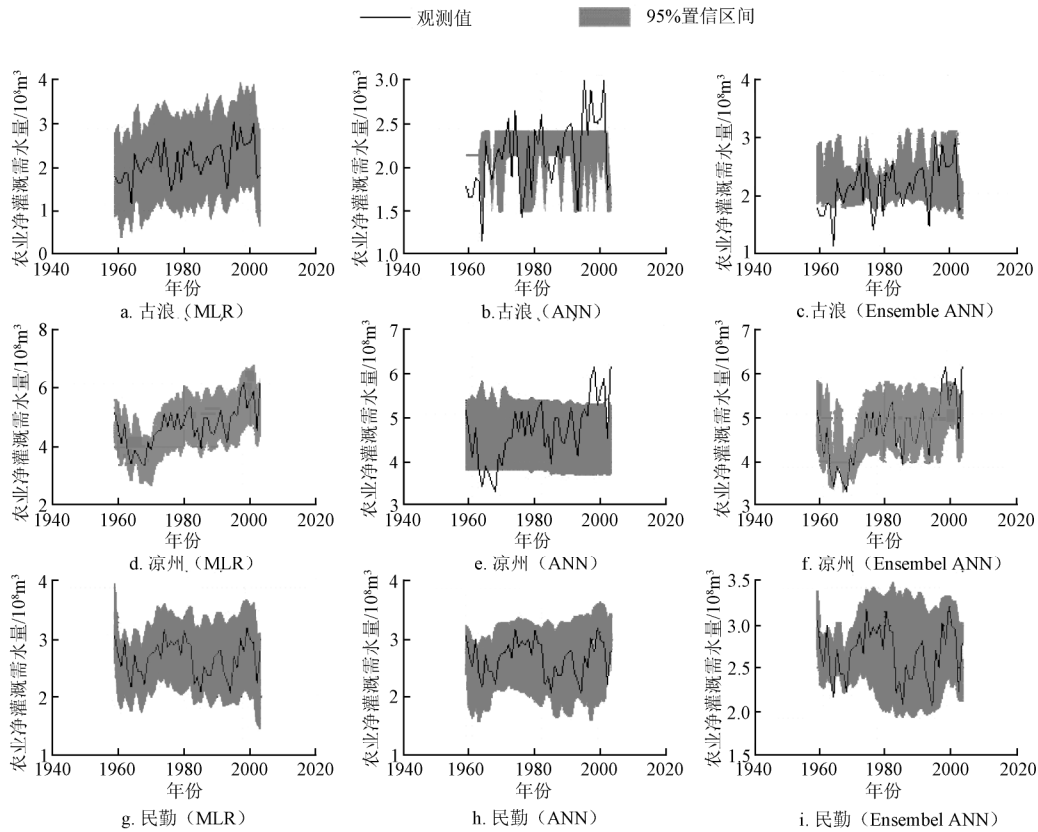


图4 古浪,凉州,民勤三种模型对输入项不确定分析的MC模拟结果

Fig.4 MC simulation results for the input uncertainty analysis in Gulang, Liangzhou, Minqin and Shiyang river basin

详细的数据情况由3种模型的不确定性分析结果图4所示,图中阴影的面积为蒙特卡洛(MC)模拟输出数据的分布(95%置信区间);图中的曲线为农业净灌溉需水量观测值。其中,由MLR对3个地区的MC模拟数值分布来看,凉州效果较好,古浪与民勤地区的效果较差;对ANN模型来说,在古浪与凉州地区均出现不合理与无趋势的变化,民勤地区的模拟数值趋势及分布区域分布效果均不好;在Ensemble ANN模型中,凉州的输出数值分布范围最窄,古浪地区部分数据未能反映观测值的浮动趋势且数值的分布范围比MLR小,民勤呈现两端的数据分布范围窄,中间的数据分布较宽的变化。在对凉州的模拟中MLR模型与Ensemble ANN模型则较为接近观测值的变化情况。

结合2.1中的模型的模拟误差对比以及不确定性分析的结果对模型进行评价:3种模型比较而言,Ensemble ANN的模拟效果优于神经网络ANN模型与MLR模型;在不确定性分析中,Ensemble ANN模型更能合理地指示输出项随输入项的变化趋势。

下游民勤的误差最小,其在不确定性分析中的不确

定性最高(d-factor值最大);中游凉州区,3种模型模拟的误差较大,不确定性分析中模型的不确定性最低(d-factor值最小)。可以认为民勤地区的农业净灌溉需水量对于人口,耕地面积,粮食产量,气象因素的变化更加敏感;对于凉州区而言,由于上游来水的优先供给以及较大的耕地面积提供了一种相对稳定的农业用水环境,农业用水相对稳定,从而农业净灌溉需水量对于外部环境因子的变化不明显。

3 结论

本文结合石羊河流域上中下游不同的农业用水情况,对古浪,凉州,民勤3个区域分别建立了净灌溉用水的多元线性回归(MLR),神经网络模型(ANN)与神经网络集成模型(Ensemble ANN),并进行了误差比较与不确定性分析。结果表明,在石羊河流域农业净灌溉需水量的经验模型建立中Ensemble ANN模型具有模拟效果最好(误差小)与不确定性表达合理的双重优势。本文中建立Ensemble ANN模型以及采用的不确定性

分析方法对于建立区域农业净灌溉需水量经验模型及其它推广应用具有一定的借鉴意义。

[参 考 文 献]

- [1] 陈玉民, 郭国双, 王广兴, 等. 中国主要作物需水量与灌溉[M]. 北京: 水利电力出版社, 1995.
- [2] Deng X P, Shan L, Zhang H, et al. Improving agricultural water use efficiency in arid and semiarid areas of China[J]. *Agricultural Water Management*, 2006, 80(1/3): 23–40.
- [3] Kang S, Su X, Tong L, et al. The impacts of human activities on the water–land environment of the Shiyang River basin, an arid region in northwest China[J]. *Hydrological Sciences journal*, 2004, 49(3): 1–427.
- [4] Leenhardt D, Trouvat J L, Gonzalès G, et al. Estimating irrigation demand for water management on a regional scale: I. ADEAUMIS, a simulation platform based on bio-decisional modelling and spatial information[J]. *Agricultural Water Management*, 2004, 68(3): 207–232.
- [5] Vörösmarty C J, Green P, Salisbury J, et al. Global water resources: vulnerability from climate change and population growth[J]. *Science*, 2000, 289(5477): 284–288.
- [6] Döll P. Impact of climate change and variability on irrigation requirements: a global perspective[J]. *Climatic Change*, 2002, 54(3): 269–293.
- [7] Gómez - Limón J A, Riesgo L. Irrigation water pricing: differential impacts on irrigated farms[J]. *Agricultural Economics*, 2004, 31(1): 47–66.
- [8] Sousa S, Martins F G, Alvim-Ferraz M, et al. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations[J]. *Environmental Modelling and Software*, 2007, 22(1): 97–103.
- [9] Ramos Nino M E, Ramirez Rodriguez C A, Clifford M N, et al. A comparison of quantitative structure activity relationships for the effect of benzoic and cinnamic acids on *Listeria monocytogenes* using multiple linear regression, artificial neural network and fuzzy systems[J]. *Journal of Applied Microbiology*, 1997, 82(2): 168–176.
- [10] Perrone M P, Cooper L N. When network architectures: equilibrium systems that pay attention[C]// *Artificial Neural Networks for Speech and Vision*. Chapman and Hall, New York, 1993: 126–142.
- [11] Optize D, Shaclik J. Actively searching for an effective neural network ensemble[J]. *Connection Science*, 1996, 8(3/4): 337–353.
- [12] Krajewski W F, Lakshmi V, Georgakakos K P, et al. A Monte Carlo study of rainfall sampling effect on a distributed catchment model[J]. *Water Resources Research*, 1991, 27(1): 119–128.
- [13] Wagner B J, Gorelick S M. Optimal groundwater quality management under parameter uncertainty[J]. *Water Resources Research*, 1987, 23(7): 1162–1174.
- [14] Kuczera G, Parent E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm[J]. *Journal of Hydrology*, 1998, 211(1/4): 69–85.
- [15] 甘肃省水利厅, 甘肃省发展与改革委员会. 石羊河流域重点治理规划[M].甘肃: 兰州, 2007.
- [16] Tong L, Kang S, Zhang L. Temporal and spatial variations of evapotranspiration for spring wheat in the Shiyang river basin in northwest China[J]. *Agricultural Water Management*, 2007, 87(3), 241–250.
- [17] Allen R G, Pereira L S. Crop Evapotranspiration - Guidelines for computing crop water requirements – FAO Irrigation and Drainage[M]. FAO. Rome, Italy, 1998, 45–46.
- [18] Mays D C, Faybishenko B A, Finsterle, S. Information entropy to measure temporal and spatial complexity of unsaturated flow in heterogeneous media[J]. *Water Resources Research*, 2002, 38(12): 1313–1315.
- [19] Fraser A M, Swinney H L. Independent coordinates for strange attractors from mutual information[J]. *Physical Review Letters*, 1986, 33(2): 1134–1140.
- [20] Zheng G L, Billing S A. Radial basis function network configuration using mutual information and the orthogonal least squares algorithm[J]. *Neural Networks*, 1996, 9(9): 1619–1639.
- [21] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123–140.
- [22] PaÁdraig C, Carney J, Jacob S. Stability problems with artificial neural networks and the ensemble solution[J]. *Artificial Intelligence in Medicine*, 2008, (20): 217–225.
- [23] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//*Proceedings of the 14th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1995, (2): 1137–1145.
- [24] Boyce M S, Vernier P R, Nielsen S E, et al. Evaluating resource selection functions[J]. *Ecological Modelling*, 2002, 157(2/3), 281–300.
- [25] Huang C M, Lee Y J, Lin D K J, Huang S Y(2007). Model selection for support vector machines via uniform design[J]. *Computational Statistics and Data Analysis*, 2007, 52(1), 335–346.
- [26] Abbaspour K C, Van Genuchten M T, Schulin R, et al. A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters[J]. *Water Resources Research*, 1997, 33(8): 1879–1892.
- [27] Talebizadeh M, Moridnejad A. Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models[J]. *Expert Systems with Applications*, 2011, 38(4): 4126–4135.
- [28] Luntz A, Brailovsky V. On estimation of characters obtained in statistical procedure of recognition[J]. *Technicheskaya Kibernetica*, 1969, 3(6): 6–12. (in Russian)

Simulation and uncertainty analysis of net irrigation requirement in agricultural area

Ma Lihua¹, Kang Shaozhong^{2*}, Su Xiaoling¹, Tongling²

(1. College of Water Resources and Architecture Engineering, Northwest Agricultural and Forest University, Yangling 712100, China;

2. Center for Agricultural Water Research in China, China Agricultural University, Beijing 100083, China)

Abstract: Net irrigation requirement is the major component of agricultural water use. In this study, the regional net irrigation requirement of agricultural area in Shiyang river basin in semi-arid region of Northwest China was analyzed. Based on the zones of local agricultural water use, the key factors are chosen by driving factor analysis (1959-2005), and then regional net irrigation requirement was simulated using multiple-linear regression (MLR), artificial neural network (ANN) and Ensemble ANN models. Then discrimination time-series Monte-Carlo (MC) simulation is used for inputs uncertainty analysis of these models. Results suggest that compared with MLR and ANN, the Ensemble ANN model show better ability in the simulation of regional net irrigation requirement with smallest error and lowest uncertainty index.

Key words: irrigation, uncertainty analysis, Monte Carlo methods, multiple-linear regression, ensemble artificial neural networks