



湖北农业科学  
Hubei Agricultural Sciences  
ISSN 0439-8114, CN 42-1255/S

## 《湖北农业科学》网络首发论文

题目: 基于随机森林算法的湖北省水资源承载力评价模型及其应用  
作者: 陆佳慧  
DOI: 10.14088/j.cnki.issn0439-8114.2020.13.017  
收稿日期: 2020-04-26  
网络首发日期: 2020-08-31  
引用格式: 陆佳慧. 基于随机森林算法的湖北省水资源承载力评价模型及其应用  
[J/OL]. 湖北农业科学. <https://doi.org/10.14088/j.cnki.issn0439-8114.2020.13.017>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

陆佳慧. 基于随机森林算法的湖北省水资源承载力评价模型及其应用[J]. 湖北农业科学, 2020, 59(13): 72-76.

# 基于随机森林算法的湖北省水资源承载力评价模型及其应用


陆佳慧  
(河海大学商学院, 南京 211100)

**摘要:**依据 PSR 框架建立水资源承载力评价指标体系,以指标阈值为依据构造随机样本,在此基础上应用随机森林算法对湖北省水资源承载力进行评价。结果表明,随机森林能够很好地拟合指标因子与水资源承载力等级之间的关系,训练模型精度高,方案可操作性强,具有一定应用性;随机森林算法能无偏识别变量重要性,其中需水模数与供水模数是湖北省水资源承载力的关键影响因素,应予以重点防控;湖北省水资源承载力空间分布不均,鄂西地区和咸宁市水资源承载力为良好级,鄂东和鄂中地区为一般等级,以武汉市为核心的江汉平原地区属较差级,建议加强水资源分区管理,以促进水资源涵养能力与经济社会同步发展。

**关键词:**随机森林算法;水资源承载力;指标体系;湖北省

中图分类号:TV211.1      文献标识码:A

文章编号:0439-8114(2020)13-0072-05

DOI:10.14088/j.cnki.issn0439-8114.2020.13.017      开放科学(资源服务)标识码(OSID): 

## A assessment model of water resources carrying capacity based on random forest algorithm and its application in Hubei province

LU Jia-hui  
(Business School of Hohai University, Nanjing 211100, China)

**Abstract:** An assessment system of water resources carrying capacity indicator was constructed according to pressure-state-response (PSR) framework, and the random samples were constructed based on index threshold. On this basis, the random forest algorithm was applied to evaluate the water resources carrying capacity in Hubei province. The results showed that the relationship between the index factors and the water resource carrying capacity could be well fitted by the random forest with a high accuracy. The scheme was of good maneuverability and certain application value. The random forest algorithm can recognize the importance of variable unbiased, and the research showed that the modulus of water demand and the modulus of water supply were the key factors affecting the carrying capacity of water resources in Hubei province. The spatial distribution of water resources carrying capacity in Hubei province was uneven. The carrying capacity of water resources in western Hubei and Xianning areas was of good grade, while that of eastern Hubei and middle Hubei was generalgrade. The Jiangnan plain area with Wuhan as the core belonged to poorgrade. It was suggested that regional management of water resources should be strengthened to promote the development of water resources conservation ability and economic society synchronously.

**Key words:** random forest algorithm; water resources carrying capacity; indicator system; Hubei province

水是人与自然赖以存续的基质,在全球气候变化与人类文明迅速发展的背景下,水资源问题较多且形势严峻,成为影响区域乃至全球可持续发展的限制因素之一。自然条件下水资源量及质量分布受

水循环过程与环境梯度影响产生时空变异,加之人居生产生活对水资源攫取过度与不合理开发造成了水质退化、水环境恶化、水资源短缺和水体污染逐渐扩展,严重影响了经济社会与环境协调、平衡演替。

厘清区域水资源本底概况及其供应能力,为经济布局、人居规划、资源利用提供信息基础,也是当前水生态文明建设的任务之一。

对于水资源承载力的量化研究,学者们主要采用指标提取-综合评判的方法,相应地提出了多元统计、系统动力、模糊数学和神经网络等评判方案<sup>[1]</sup>。需要指出的是,常规方法普遍存在不足,一是指标数据噪声或专家估值易造成赋权和评价分级可靠性的争议,二是指标间共线性使指标体系信效度降低而影响了评价的准确性<sup>[1,2]</sup>。为此,提出应用随机森林(Random forest)机器学习算法。该方法避开了从归纳到演绎的传统过程,对噪声数据具有较好的鲁棒性且对多元线性关系不敏感,还能有效防止过拟合<sup>[3]</sup>,作为人工智能的优秀算法之一,其在模式识别、概率测算、函数逼近和回归等问题应用中取得了良好的效果<sup>[4]</sup>。鉴于此,以湖北省为案例区,旨在分析探讨基于随机森林算法的水资源承载力评价流程,以为省域水资源综合评判和空间优化管理提供参考依据。

## 1 数据源与研究方法

### 1.1 研究区与数据源

湖北省地处华中亚热带季风气候区,区域降水量丰富(800~1 600 mm),河川径流长度达 5.92 万 km,水资源总量达 825.28 亿 m<sup>3</sup>(2011 年)。全省河湖众多而水资源时空分布不均,时间上集中于夏、秋季节,空间主要分布于鄂西、鄂东山地;江汉平原地区存在水质污染等问题。

以湖北省的 17 个地级单位为评价对象,依据水

资源承载力评价指标体系(表 1),从《中国统计年鉴》《湖北统计年鉴》中提取相关指标数据,考虑到数据的平稳和现势性,取 2011—2015 年数据的平均值。

### 1.2 水资源承载力评价指标体系分级标准

基于水生态系统的完整性选取可量化的指标,进而实施综合评判,是当前水资源承载力研究的主流方法之一。水资源承载力客体具有结构性、要素关联性等特点,指标因子不仅要能体现区域资源禀赋和生态要素现状,还要能反映人类扰动和自然环境过程与水资源数量、水环境容量之间的互馈。目前有关水资源承载力的指标难以统一,其中 PSR(Pressure-State-Response)框架以决策者思维阐释水资源系统整体性、动态性和多维性<sup>[5]</sup>,对生态评价模型中的指标确定具有普适意义。遂以该框架将水资源承载力逐层划分为压力、状态、响应维度,结合区域现状,构建了湖北省的水资源承载力评价指标体系(表 1)。该方法以系统动力学和分层理念定量描述水资源-经济社会-生态环境三大子系统的内在联系。

分级标准是对水资源承载力进行准确评价的关键,指标既要能反映水资源承载力的渐进演替规律,还应具有公开性、统一性。虽然国内外学者展开了大量研究,但出于水资源承载力区域性的特点,尚未形成具有通用性的阈值范式。鉴于此,以《湖北省工业与生活用水定额》、国际水安全标准<sup>[6]</sup>以及其他学者研究经验<sup>[1,2]</sup>,确定了湖北省水资源承载力指标分级阈值(表 1)。将其划分为 5 个等级:理想(I),表征水环境优越,能够提供充分承载能力;良好(II),

表 1 区域水资源承载力评价指标体系与分级标准

| 维度 | 指标层   | 计算方法         | 性质 | 评价分级   |               |               |             |      |
|----|---|--------------|----|--------|---------------|---------------|-------------|------|
|    |   |              |    | I      | II            | III           | IV          | V    |
| 压力 | 人均水资源量(A1)//m <sup>3</sup>                  | 水资源总量/总人口    | +  | >3 000 | (2 000,3 000] | (1 000,2 000] | (500,1 000] | ≤500 |
|    | 需水模数(A2)//万 m <sup>3</sup> /km <sup>2</sup> | 需水量/总面积      | -  | <1     | [1,3)         | [3,10)        | [10,15)     | ≥15  |
|    | 水资源开发利用率(A3)//%                             | 供水总量/水资源总量   | -  | <10    | [10,20)       | [20,40)       | [40,60)     | ≥60  |
|    | 万元 GDP 耗水量(A4)//m <sup>3</sup>              | 用水总量/GDP     | -  | <20    | [20,50)       | [50,90)       | [90,120)    | ≥120 |
|    | 日生活用水定额(A5)//m <sup>3</sup>                 |              | -  | <50    | [50,70)       | [70,130)      | [130,170)   | ≥170 |
|    | 万元 GDP 排污量(A6)//m <sup>3</sup>              | GDP/污水总量     | -  | <10    | [10,20)       | [20,30)       | [30,40)     | ≥40  |
| 状态 | 产水模数(A7)//万 m <sup>3</sup> /km <sup>2</sup> | 水资源总量/区域面积   | +  | ≥100   | (70,100]      | (30,70]       | (10,30]     | ≤10  |
|    | 供水模数(A8)//万 m <sup>3</sup> /km <sup>2</sup> | 供水量/区域面积     | +  | ≥35    | (15,35]       | (5,15]        | (1,5]       | ≤1   |
|    | 森林覆盖率(A9)//%                                | 森林面积/总面积     | +  | ≥55    | (40,55]       | (25,40]       | (10,25]     | ≤10  |
|    | 降水量(A10)//mm                                |              | +  | ≥2 000 | (1 400,2 000] | (800,1 400]   | (400,800]   | ≤400 |
| 响应 | 环境投资占比(A11)//%                              | 环境治理投资总额/GDP | +  | ≥5     | (4,5]         | (2,4]         | (1,2]       | ≤1   |
|    | 污水处理率(A12)//%                               | 污水处理量/排污总量   | +  | ≥99    | (90,99]       | (80,90]       | (65,80]     | ≤65  |
|    | 节水灌溉率(A13)//%                               | 节水灌溉面积/灌溉面积  | +  | ≥80    | (60,80]       | (40,60]       | (20,40]     | ≤20  |

注:正向指标(+)值越大或负向指标(-)越小,表明水资源承载力越好;反之水资源承载力越差



表示水环境系统受轻微破坏,易于修复,水资源状况乐观、开发潜力大;一般(Ⅲ),表示水环境系统受到破坏,质量脆弱,尚可维持基本服务功能;较差(Ⅳ),表示水环境系统质量堪忧、支撑能力明显退化,只能维持部分功能;极差(Ⅴ),表示其严重退化,恢复非常困难,难以提供生态承载与社会服务功能。

1.3 基于随机森林的评价模型设计

应用随机森林算法实施水资源承载力评价,关键在于依据水资源承载力分级标准构建其与指标系统中的单一指标间的模式规则,在 Rstudio 环境下其具体流程如下。

Step 1:依据评价标准,应用 R 软件中的 runif 函数在各等级标准区间随机生成 200 组样本数据,共计样本 1 000 组;对于样本期望输出,以 1、2、3、4、5 分别表示水资源承载力等级。

Step 2:从 1 000 组样本中随机选取 700 组作为训练样本,另外 300 组为独立验证样本;13 个指标数据为输入向量,5 个期望值为输出向量,通过 Random forest 函数进行网络训练。

Step 3:算法中的出包错误率(Out-of-bag error, OOB error)为模型精度的衡量,对于敏感参数 mtry 和 ntree 的选择,通过网格搜索法观察 OOB error 随参数的变化进行优选。如图 1 所示,当 mtry 为 4 时,OOB error 最小;当 ntree 大于 600 时,OOB error 较小而趋于稳定,遂将其确定为最优参数。

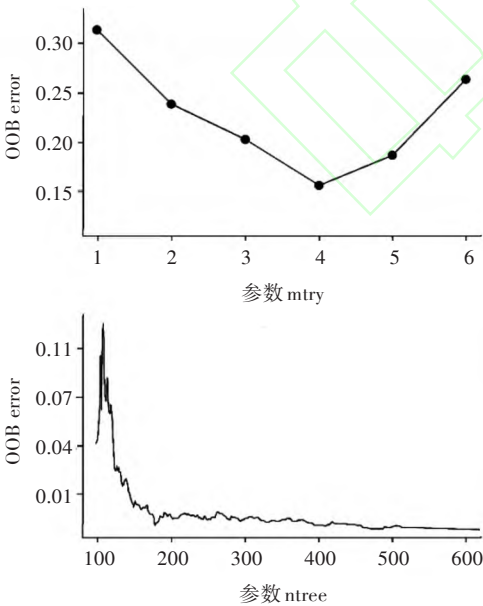


图1 模型精度随参数的变化

Step 4:以湖北省 17 个市水资源承载力指标原始数据为输入变量,将训练好的模型代入并拟合,输出各市水资源承载力分级值,并按四舍五入法归类<sup>[7]</sup>,以评价水资源承载力等级。

Step 5:为了解算法可适性,采用决定系数( $R^2$ )、均方根误差(RMSE)、平均绝对误差(MSE)对训练样本和独立样本的拟合结果进行评价<sup>[8]</sup>。通常, $R^2$ 接近于 1,RMSE、MSE 接近于 0 时,表明算法拟合度高。由表 2 可知,训练样本的  $R^2$  为 0.996,RMSE 与 MSE 均较小,验证样本  $R^2$  为 0.975,RMSE、MSE 为 0.074、0.015,表明该算法可靠,可用于对目标样本的预测。训练模型存在一定误差,将其代入验证样本进行测试时,由于误差传递而精度降低,故验证样本的精度略小于训练样本。但总体来看,基于随机森林算法的水资源承载力评价模型具有一定应用性。

表 2 随机森林算法拟合结果

| 样本   | $R^2$ | RMSE  | MSE   |
|------|-------|-------|-------|
| 训练样本 | 0.996 | 0.023 | 0.006 |
| 验证样本 | 0.975 | 0.074 | 0.015 |

2 结果与分析

2.1 湖北省水资源承载力评价结果

基于前述方法,对湖北省 17 个市水资源承载力进行预测,将输出值归类后于 ArcGIS10.3 平台上进行空间可视化,如图 2 所示。湖北省水资源承载力呈现一定空间聚集性,鄂西地区的神农架、恩施、十堰、宜昌等地和咸宁属于良好级别(Ⅱ),表明该地区水资源涵养能力与经济社会发展结合较好,水资源开发潜力大。鄂西地区自然生态环境较好,河流发育密集,水资源蕴藏丰富,而该区人口分布稀疏、产业发展滞后,故用水压力较小,水资源供应充足。江汉平原的武汉、潜江、仙桃、天门、孝感等地的水资源承载力为较差(Ⅳ)级别,这些地方水环境受到一定破坏,水资源可再生能力难以维继经济社会发展。这是由于该区人均国土空间狭小、地表产流利用相对不足,同时作为省域经济核心、人口密集区,区域用水需求量大,因而水资源开发潜力、供需状况不乐

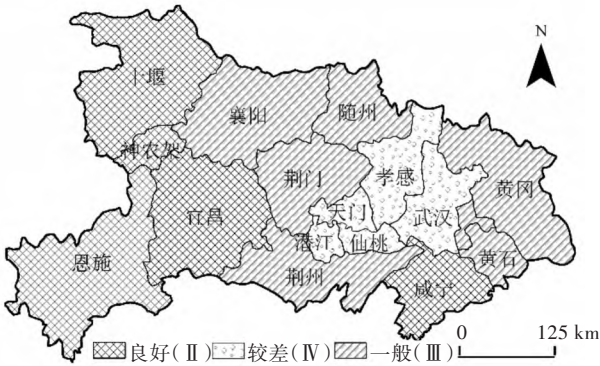


图2 湖北省水资源承载力空间分布

观。鄂东(黄冈、黄石)和鄂中(襄阳、荆门、随州、荆州)地区的水资源承载力为一般水平(Ⅲ),表明区域水资源基本能够支撑经济社会发展,但承载力脆弱,对此应加强水资源优化配置与水环境保护。

2.2 水资源承载力指标边际效应分析

图 3 为水资源承载力与各指标因子间的边际效应,图中横坐标为影响因素的变化范围,纵坐标为边际效应函数( $f$ )。由图 3 可知,水资源承载力与其影响因子之间并非简单的线性关系,随着指标数值变化,水资源承载力对其条件期望呈阶梯状变化。边际效应函数变化趋势表明,水资源承载力与需水模数(A2)、水资源开发利用率(A3)、万元 GDP 耗水量(A4)、日生活用水定额(A5)和万元 GDP 排污量(A6) 5 项因子呈负相关,与其他因子呈正相关,这与表 1 中各指标属性描述相一致。边际函数在不同区间的线性斜率反映了水资源承载力影响因子期望值的大

小,有助于理解指标因子对水资源承载力的影响。

2.3 水资源影响因素分析

随机森林模型基于 bootstrap 抽样生成决策树的算法克服了多重共线性、数据噪声的影响,能良好规避水资源承载力评价指标因子间的相关性或不符合正态分布等问题,因而对影响因子重要性的估计是无偏的<sup>[4]</sup>。图 4 为水资源承载力对各指标的 MDG 重要性分值。可知,供水模数(A8)和需水模数(A2)的 MDG 值最大,分别为 0.686、0.657,表明其是湖北省水资源承载力的关键影响因子;万元 GDP 排污量(A6)、日生活用水定额(A5)、水资源开发利用率(A3)、森林覆盖率(A9)、人均水资源量(A1)、污水处理率(A12)和产水模数(A7)的 MDG 值次之,介于 0.602~0.644,对水资源承载力有较大影响;而其他因子的 MDG 值较小。

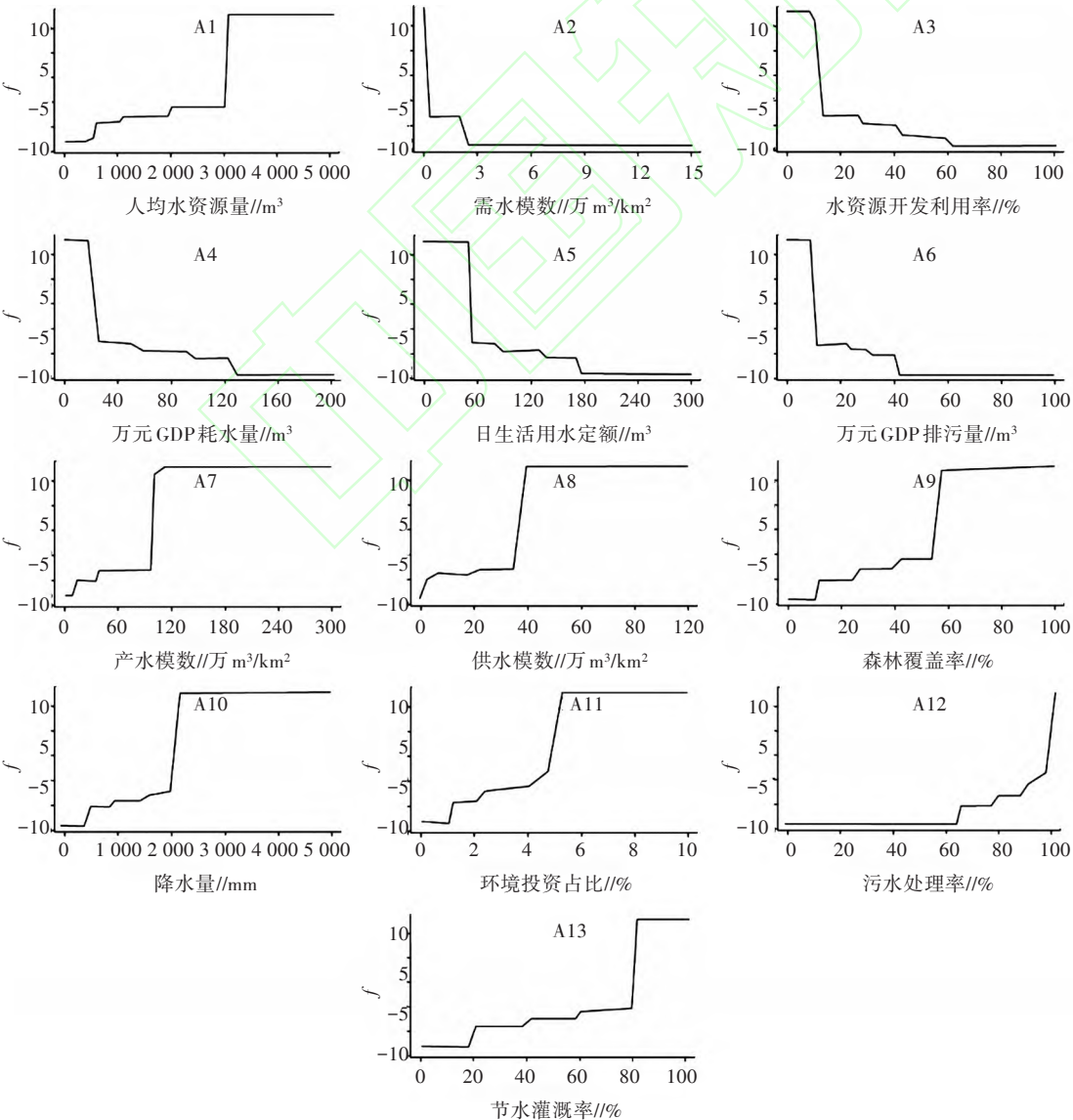


图3 各指标因子对水资源承载力的影响

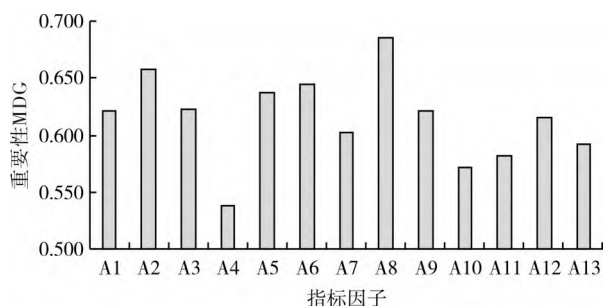


图4 指标因子的重要性

### 3 结论

依据水资源与经济社会、环境之间的系统动力反馈,选出 13 项指标因子构建水资源承载力评价指标体系,尝试将随机森林算法应用于湖北省水资源承载力评价中,并详述了随机森林原理及其应用过程,得出结论如下。

1)随机森林算法能够良好地拟合水资源承载力指标数据与水资源承载力相应等级值之间的规律,从而实现智能评价;该算法结构简易、调参量少、拟合精度高,在 R 平台上可操作性强,具有一定应用意义。

2)随机森林以组合树算法深入挖掘数据内部之间多维复杂关系,从高维特征空间识别与水资源承载力密切相关的主要变量,实现对影响因素重要性

的无偏估计。研究表明,供水模数和需水模数是湖北省水资源承载力的关键影响因子。

3)湖北省水资源承载力呈空间聚集性特征,鄂西地区和咸宁水资源承载力最好,为良好等级;鄂东和鄂中地区处于一般水平;以武汉为中心的江汉平原地区水资源承载力属较差级别。鉴于此,应当加强省域水资源分区管理和科学规划。

#### 参考文献:

- [1] 王建华,翟正丽,桑学锋,等. 水资源承载力指标体系及评判准则研究[J]. 水利学报,2017,48(9):1023-1029.
- [2] 朱一中,夏 军,谈 戈. 关于水资源承载力理论与方法的研究[J]. 地理科学进展,2002,21(2):180-188.
- [3] BREIMAN L. Random forests[J]. Machine learning,2001,45(1):5-32.
- [4] 方匡南,吴见彬,朱建平,等. 随机森林方法研究综述[J]. 统计与信息论坛,2011,26(3):32-38.
- [5] WOLFSLEHNER B, VACIK H. Evaluating sustainable forest management strategies with the Analytic Network Process in a Pressure-State-Response framework [J]. Journal of environmental management, 2008, 88(1):1-10.
- [6] 徐元凤. ISO 发布关于水安全的国际专题组协议[J]. 中国标准化,2008(6):78-78.
- [7] 吴 敏,温小虎,冯 起,等. 基于随机森林模型干旱绿洲张掖盆地地下水水质评价[J]. 中国沙漠,2018,38(3):1-7.
- [8] SABATIA C O, BURKHART H E. Predicting site index of plantation loblolly pine from biophysical variables [J]. Forest ecology & management, 2014, 326:142-156.
- [9] 气象,2004,22(2):57-61.
- [10] 董大学,钟良平,李玉山. 冬小麦产量与底墒关系分析[J]. 水土保持通报,1993(5):58-61.
- [11] 李 超,刘文兆,林 文. 黄土塬区冬小麦产量及水分利用效率对播前底墒变化与生育期差别供水的响应[J]. 中国农业科学,2017,50(18):3549-3560.
- [12] 刘长海,刘亚梅,于晓春. 2016年夏季“伏旱”对九三垦区玉米生长发育的影响[J]. 现代化农业,2017(11):15-17.
- [13] 王艳波,徐 峰,赵鹏飞. 北方农区伏旱现状及对策[J]. 现代农业科技,2016(17):201-203.
- [14] 陈思宇,张杰丹. 甘肃省广河县伏旱特征分析及防御对策[J]. 北京农业,2016(2):157-159.
- [15] 李 梗,刘晓冉,刘 德. 重庆市伏旱时空变化特征[J]. 气象科技,2011,39(1):27-32.
- [16] 韩世刚,苟 思,吉 莉. 重庆市伏旱频次气候特征分析[J]. 安徽农业科学,2010,38(29):16372-16373.
- [17] 隋 东,梁 红,马福全. 2009年沈阳地区高温伏旱成因分析[J]. 安徽农业科学,2010,38(9):4693-4695.
- [18] 陈立娟,王瑞华. 伏旱对玉米生长的影响及防御对策[J]. 安徽农学通报,2015(7):57-58.
- [19] 方建刚,白爱娟,肖科丽. 陕西伏旱气候特征及成因分析[J]. 干旱地区农业研究,2009,27(2):28-34.
- [20] 王帮能,徐世海,谭云廷. 丰都县伏旱时空分布特征及防御对策[J]. 贵州气象,2014,38(2):41-43.
- [21] 田 红,徐 敏,李 树. 安徽省伏旱特征及预报研究[J]. 高原气象,2006,25(4):731-736.
- [22] 李 君,陈秀杰,司志华. 山东伏期降水时空的分布特征[J]. 安徽农业科学,2009,37(22):10565-10568.
- [23] 钟继祥. 广安伏旱及防御对策浅析[J]. 四川气象,2003,23(1):34-41.
- [24] 赵 森,蔡祖炎. 盆东伏旱气候的时空分布及其形成原因[J]. 四川气象,2000,20(4):5-8.
- [25] 唐红艳. 伏旱对兴安盟粮食作物产量的影响及对策[J]. 内蒙古气象,2001(2):37-39.
- [26] 《达州伏旱灾害对农业生产影响评估研究》课题评审专家组. 达州伏旱灾害对农业生产影响(预)评估鉴定意见[J]. 四川气象,2003(4):59.
- [27] 薛景轩,赵天虎. 冬小麦收墒期降水量的分析及预报[J]. 甘肃农村科技,1997(4):22-23.
- [28] 贾秀领,马瑞昆,张全国. 近 20 年冬小麦供水量与产量关系变化分析[J]. 华北农学报,2009,24(Z1):214-217.
- [29] 张云兰. 高温伏旱区旱地农作系统水分供需平衡特征与生态适应性研究[D]. 重庆:西南大学,2010.
- [30] 刘富明. 盆东伏旱气候规律研究及发展农业生产的对策建议[J]. 四川气象,2000,20(4):2-4.
- [31] 张存杰,董安祥,白虎志,等. 甘肃省河东地区伏旱的小波分析[J]. 应用气象学报,1998(3):36-42.
- [32] 阎凤台,郭江勇. 陇东伏旱分析及其预报[J]. 甘肃气象,1990(3):2-4.
- [33] 冉敬才. 川东南山区玉米避伏旱技术及其应用[J]. 作物杂志,1994(2):27.

(上接第 42 页)