

分类号:
学号: 20212010029

密级: 公开
单位代码: 10759

石河子大学 硕士学位论文



基于机器学习的膜下滴灌棉花灌溉决策 优化研究

学位申请人	曹缘
指导教师	王振华 教授
申请学位门类级别	工学硕士
学科、专业名称	农业工程
研究方向	农业水土工程
所在学院	水利建筑工程学院

中国·新疆·石河子
2024年5月

分类号：
学号：20212010029

密级：公开
单位代码：10759

石河子大学

硕士学位论文



基于机器学习的膜下滴灌棉花灌溉决策 优化研究

学位申请人	曹缘
指导教师	王振华 教授
申请学位门类级别	工学硕士
学科、专业名称	农业工程
研究方向	农业水土工程
所在学院	水利建筑工程学院

中国·新疆·石河子

2024年5月

**Optimization of cotton irrigation decision for drip irrigation under
film based on machine learning**

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Engineering

By

Cao Yuan

(Agricultural Engineering)

Dissertation Supervisor: Prof. Wang Zhen-hua

May, 2024

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所呈交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：曹缘

时间：2024年5月5日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：曹缘

时间：2024年5月5日

导师签名：

时间：2024年5月5日

曹缘

摘要

目的:棉花是新疆的主要种植作物之一,种植用水过程中最主要的就是灌溉,作物需水量是生育期内植株蒸腾、棵间蒸发和作物体内的水量,是确定作物灌溉用水量的重要部分。影响膜下滴灌棉花需水量的因素具有高维、非线性的特点,目前部分学者通过构建算法模型来预测作物需水量,但是应用于膜下滴灌棉花的研究较少,且这些预测模型大多聚焦于大时空范围蒸散,对实际需水量的研究仍有较大空间。因此,本研究旨在探究不同环境因素对膜下滴灌棉花需水量的影响关系,设计出一种预测精度好、性能优良的膜下滴灌棉花需水量预测方法,探究适合需水量预测的最佳环境参数,并进行适用性评价和模型对比分析,为膜下滴灌棉花灌溉提供决策参考。

方法:研究采用相关性分析方法量化各环境因素对膜下滴灌棉花需水量的影响关系和影响程度,并构建输入参数组合和需水预测模型,以预测精度和预测代价为选取准则,明确适合预测模型的最佳输入参数,验证预测模型的可行性。在此基础上探究在不同时间尺度下的模拟效果,验证需水预测模型的适用性。结合前人对预测模型的研究构建基于不同优化算法的机器学习模型,以 R^2 、RMSE、MAE、MAPE 为评价指标对比分析验证预测模型的有效性。结合需水预测管理系统提供决策支持和数据管理。

结果: (1) 在小时尺度下, Spearman 和 MIC 相关性分析都得到了一致的重要性程度,各环境因素对膜下滴灌棉花蒸散量的影响程度排序为:辐射>气温>相对湿度>风速>土壤温度>气压>土壤水分。日尺度下相关性存在差异,最主要的影响因素是气温。(2) 依据相关性结果对环境因素进行组合,在仅有辐射参数时,模型已经表现出一定的预测能力,随着输入参数的增加模型精度稳步提升,加入土壤温度后预测精度提升显著,拟合优度达到了 0.923。综合计算精度和运行代价,输入辐射、气温、相对湿度、风速和土壤温度时预测模型有最好的综合性能。在相同输入因素下模拟了日尺度的预测效果, R^2 提高了 0.86%, MAPE 降低了 57.48%,验证了需水预测模型的适用性。(3) 构建不同预测模型进行对比分析,基于智能算法优化的混合模型都表现出了更优的预测性能,在混合模型中,WOA-XGBoost 的 R^2 提升了 0.33%~2.1%,预测误差 RMSE、MAE、MAPE 分别减少了 1.54%~9.86%、2.56%~19.15%、0.92%~38.57%,WOA 算法和 XGBoost 模型独特的优势可以应用于棉花需水预测中。(4) 采用软件工程技术,设计数据库和 Web 页面,实现环境数据和业务数据的存储,开发可信息化平台,完成用户管理、数据管理、需水预测管理等功能。

结论:影响膜下滴灌棉花需水量的环境因素错综复杂,辐射是小时尺度下膜下滴灌棉花需水量的主要影响因素,日尺度下为气温。相关性中度相关以上环境因素已展现出较好的预测效果,增加因素有利于提升预测模型精度,除气象因素外,土壤因素也显著提高了预测模型性能。样本集和模型超参数是预测模型的核心,群体智能优化算法的寻优能力为模型提供了最优超参数组合,减少预测值和实测值的误差,研究提出的需水预测模型在不同尺度、不同模型中均展示出了更优的预测效果,能够为棉花需水预测提供科学合理的决策支持。

关键词:膜下滴灌;棉花;需水量;机器学习;群体智能算法

Abstract

Objective: Cotton is one of the main crops in Xinjiang. Irrigation is the most important part in the process of planting water. Crop water requirement is the water amount of plant transpiration, inter-grain evaporation and crop body during the growth period, which is an important part of determining crop irrigation water consumption. The factors affecting the water requirement of cotton under mulched drip irrigation have the characteristics of high dimension and nonlinearity. At present, some scholars predict crop water requirement by constructing algorithm models, but there are few studies on cotton under mulched drip irrigation, and most of these prediction models focus on evapotranspiration in a large space-time range, and there is still a large space for the study of actual water requirement. Therefore, the purpose of this study is to explore the influence of different environmental factors on the water demand of cotton under mulched drip irrigation, design a prediction method of cotton water demand under mulched drip irrigation with good prediction accuracy and excellent performance, explore the best environmental parameters suitable for water demand prediction, and carry out applicability evaluation and model comparison analysis to provide decision-making reference for cotton irrigation under mulched drip irrigation.

Methods: The correlation analysis method was used to quantify the influence relationship and influence degree of various environmental factors on the water demand of cotton under mulched drip irrigation, and the input parameter combination and water demand prediction model were constructed. Taking the prediction accuracy and prediction cost as the selection criteria, the optimal input parameters suitable for the prediction model were determined, and the feasibility of the prediction model was clarified. On this basis, the simulation results at different time scales are explored to verify the applicability of the water demand prediction model. Combined with the previous research on the prediction model, a machine learning model based on different optimization algorithms was constructed, and R^2 , RMSE, MAE, and MAPE were used as evaluation indicators to compare and analyze the validity of the prediction model. Combined with water demand forecasting management system to provide decision support and data management.

Results: (1) At the hourly scale, Spearman and MIC correlation analysis have obtained a consistent degree of importance. The degree of influence of various environmental factors on cotton evapotranspiration under mulched drip irrigation is ranked as follows : radiation > temperature > relative humidity > wind speed > soil temperature > air pressure > soil moisture. There are differences in the correlation at the daily scale, and the most important factor is temperature. (2) According to the correlation results, the environmental factors are combined. When there are only radiation parameters, the model has shown a certain predictive ability. With the increase of input parameters, the accuracy of the model is steadily improved. After adding soil temperature, the prediction accuracy is significantly improved, and the goodness of fit reaches 0.923. Considering the calculation accuracy and operation cost, the prediction model has the best comprehensive

performance when input radiation, air temperature, relative humidity, wind speed and soil temperature. The prediction effect of daily scale was simulated under the same input factors. R² increased by 0.86 % and MAPE decreased by 57.48 %, which verified the applicability of the water demand prediction model. (3) In the hybrid model, the R² of WOA-XGBoost is increased by 0.33 % ~ 2.1 %, and the prediction errors RMSE, MAE and MAPE are reduced by 1.54 % ~ 9.86 %, 2.56 % ~ 19.15 % and 0.92 % ~ 38.57 %, respectively. The unique advantages of WOA algorithm and XGBoost model can be applied to cotton water demand prediction. (4) Using software engineering technology, design database and Web pages, realize the storage of environmental data and business data, develop information platform, complete user management, data management, water demand forecasting management and other functions.

Conclusion: The environmental factors affecting the water requirement of cotton under mulched drip irrigation are complicated. Radiation is the main factor affecting the water requirement of cotton under mulched drip irrigation at the hourly scale, and the temperature at the daily scale. The environmental factors above moderate correlation have shown good prediction results. Adding factors is conducive to improving the accuracy of the prediction model. In addition to meteorological factors, soil factors also significantly improve the performance of the prediction model. The sample set and model hyperparameters are the core of the prediction model. The optimization ability of the swarm intelligence optimization algorithm provides the optimal hyperparameter combination for the model and reduces the error between the predicted value and the measured value. The proposed water demand prediction model shows better prediction results in different scales and different models, which can provide scientific and reasonable decision support for cotton water demand prediction.

Key words: Mulched drip irrigation; Cotton; Water requirement; Machine learning; Swarm intelligence algorithm

目录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 作物需水量测算研究现状.....	3
1.2.2 机器学习算法在作物需水量预测中的应用.....	5
1.2.3 国内外研究现状总结.....	6
1.3 研究目标及研究内容.....	6
1.3.1 研究目标.....	6
1.3.2 研究内容.....	6
1.4 技术路线.....	7
第 2 章 研究模型与方法.....	9
2.1 预测模型关键技术.....	9
2.1.1 集成学习.....	10
2.1.2 支持向量机模型.....	12
2.2 群体智能优化算法.....	15
2.2.1 鲸鱼优化算法.....	15
2.2.2 粒子群优化算法.....	17
2.3 交叉验证方法.....	19
2.4 评价指标.....	19
第 3 章 膜下滴灌棉花需水量影响因素分析.....	21
3.1 研究区概况.....	21
3.2 数据来源.....	21
3.3 数据预处理.....	22
3.4 膜下滴灌棉花影响因素相关性分析.....	22
3.4.1 影响因素 Spearman 相关性分析.....	22
3.4.2 影响因素 MIC 相关性分析.....	26
3.4.3 不同时间尺度下影响因素相关性分析.....	27
3.5 小结.....	29
第 4 章 膜下滴灌棉花需水预测模型建立及优选.....	30
4.1 环境数据预测蒸散量模型的建立.....	30

4.1.1 样本集选取与划分	30
4.1.2 输入特征组合	33
4.2 基于 WOA-XGBoost 的膜下滴灌棉花蒸散量预测模型	34
4.2.1 WOA-XGBoost 模型构建	34
4.2.2 模型参数设置	35
4.2.3 不同输入组合下模型预测及评价	36
4.2.4 日尺度下模型预测分析	41
4.3 基于不同优化算法的蒸散量预测模型	42
4.3.1 PSO 算法优化 XGBoost 模型	42
4.3.2 WOA 算法优化 SVR 模型	44
4.4 不同蒸散量预测模型对比分析	45
4.5 小结	48
第 5 章 膜下滴灌棉花需水预测管理系统	50
5.1 系统需求分析	50
5.2 系统设计概要	51
5.2.1 系统框架设计	52
5.2.2 数据库设计	52
5.3 系统开发环境与实现	53
5.4 系统界面展示	55
5.5 小结	58
第 6 章 结论与展望	59
6.1 研究结论	59
6.2 展望	60
参考文献	61
致谢	68
作者简介	69
导师评阅表	70

第1章 绪论

1.1 研究背景及意义

水资源是人类生存和现代社会经济发展必不可少的物质基础和生态系统要素，水资源不仅关系着人类、自然、生态的和谐发展，还关系着一个国家经济生产的稳定发展和生态繁荣^[1]。全球气候变化、人口数量增加和经济社会发展都在加剧水资源的紧张程度^[2]。我国水资源总量丰富，占全球可利用水资源总量的 5.1%，但是由于人口众多，致使人均水量极少^[3]。中国水资源公报显示，2022 年，全国水资源总量为 27088.1 亿 m^3 ，降水总量和水资源总量比多年平均值偏少。中国作为农业大国，农业用水需求一直处于主要地位，占总用水量的 63.0%，相比 2021 年增加了 137.0 亿 m^3 ^[4]。但是全国降水量和径流量由东南沿海向西北内陆递减，水资源分布与生产生活需求不匹配，西北干旱和半干旱地区农业用水分配和供需管理问题仍然面临挑战^[5-6]。

新疆位于亚欧内陆，处于中国西北地区，地域辽阔，区域面积占据全国总面积的六分之一，是中国面积最大的内陆省份。全年气温变化大，日照时间长，蒸发旺盛，独特的“三山夹两盆”地形阻隔了海洋气流的流入，导致降雨稀少，年均降水不足 200 mm，是典型的干旱半干旱地区^[7-9]。全疆水资源空间时间分布不均匀，呈现出由北向南、由西向东递减的趋势^[10-11]，农业灌溉用水占新疆总用水量的 90% 以上，但是灌溉水利用系数不高，远低于发达国家^[12]。随着农业发展和人口的增加，用水和供水问题依然严重。同时新疆地区长期存在土壤盐碱化问题，如果不能科学合理的利用水资源，优化灌溉方式，必然会增加土壤盐分累积，进而胁迫农田作物的生长发育^[13-14]。因此在新疆大力发展节水灌溉技术具有重大意义。

棉花是纺织产业的主要原料之一，具有耐旱、耐盐碱的特点，我国种植棉花的历史悠久，至今已发展成为世界产棉大国^[15-16]。新疆地区独特的自然资源为棉花种植提供了先决条件。2020 年，新疆棉花种植面积为 250.19 万公顷，占全国总种植面积的 78.9%，棉花产量为 516.1 万吨，占全国棉花总产量 87.3%，约占世界棉花产量 20% 以上，棉花种植持续向新疆地区聚拢^[17]。棉花种植最主要和最首要的条件就是灌溉，由于地区气候和水资源条件的局限性，衍生了膜下滴灌技术，将覆膜种植和滴灌技术相结合，将灌溉水直接滴入作物根部，不仅能够显著减少植物蒸腾和裸间蒸发的水分损失，还可以改善土壤环境，减少碱性物质对土壤的侵蚀^[18-20]。膜下滴灌技术是新疆农业可持续发展的最为有效的重要技术之一，已经广泛应用于棉田中。在灌溉方式中，主要是采用传统的灌溉方式，由人工凭借过往经验操控水泵的启停。这种方式对灌溉时间和灌溉水量没有明

确限制，缺乏科学合理的灌溉决策，从而影响灌溉效率。已有的灌溉系统大多也依靠经验知识进行，智能化水平不足^[21]。随着信息化水平的不断发展，逐渐衍生出人工智能、大数据等技术，通过分析环境数据和构建灌溉模型实现作物的种植管理，这使得现代农业灌溉逐渐向“精准控制”以及“科学灌溉”的方向发展，从而使得农业灌溉更加有效率，逐步向信息化模式转变。

作物需水量是满足生育期内植株蒸腾、棵间蒸发和作物体内所消耗的水量，在现实生活中，植株体内的水分可以忽略不计，植株蒸腾、棵间蒸发组成的蒸散量即为作物的需水量^[22]。不同的作物对水分有不同的需求，每个生长阶段对水分的需求也不尽相同，在作物生长和发育时期灌溉水的重要性尤为凸显，要保证提供足够的灌溉水量，过少会出现干旱胁迫引起作物减产甚至凋零，过量会导致用灌溉水资源浪费^[23]。所以对作物进行农业灌溉时要根据作物本身的需水量、需水规律进行灌溉管理。膜下滴灌作为一种区域典型的节水技术，通过对土壤和作物根部进行覆膜预防水分蒸发，一定程度上改变了作物的生长环境，所以蒸散过程和蒸散特征也相应的发生了改变^[24]。现有的作物蒸散模型就与膜下滴灌不匹配，这就需要对膜下滴灌棉花的影响特征进行分析，研究膜下滴灌棉花在生长发育阶段内的需水状况，作为灌溉过程中的决策方法，为灌溉用水提供科学合理的理论支撑。

目前国内外学者在作物蒸散量计算方面主要有基于仪器测量、基于公式计算、基于机器学习模型等方法。影响棉花需水量的因素具有高维、非线性的特点，基于模型的方法为非线性问题提供了新的解决方案，具有较好的收敛性、实时性和一定的泛化能力，近几年在优化灌溉决策中得到了广泛应用。但这些模型在预测精度、复杂度及模拟速度等仍有不足，与灌溉控制平台结合程度不高。研究发现，模型的算法选择和结构参数权重对其性能有直接影响，因此设计出一种适用于膜下滴灌棉花的需水预测模型，为棉花灌溉提供决策依据，并结合可视化平台进行数据管理，对优化灌溉制度，提高水资源利用率具有重要的意义。

1.2 国内外研究现状

研究和发展节水灌溉技术是目前农业灌溉发展中的必由之路，而灌溉的核心是将水分直接输送到作物根系进行吸收。蒸发蒸腾是衡量作物需水量的一个重要指标，以土壤蒸发和植物蒸腾的形式转化为水蒸气向大气的运动，其值的准确度对作物的需水量有重要影响。因此，有必要对蒸散发进行适当的评价，指导灌溉决策防止灌溉过度或不足。

1.2.1 作物需水量测算研究现状

1.2.1.1 仪器测量

蒸散是液态水转化为水蒸气的过程,需要一定的能量供给,常见的能量有太阳辐射、温度、风速、水汽压等^[25]。在蒸散量实测方面,已经有很多专业仪器可以应用于测算从而计算出蒸散量。目前常用的测定方法有波文比-能量平衡法、涡度相关法、蒸渗仪法、遥感法等。

涡度相关法基于微气象学原理,利用下垫面显热和潜热的湍流脉动值计算协方差获得蒸散量,是一种可以连续且长时间监测的方法^[26]。Tie Q 等^[27]研究了华北亚湿润山地森林年、月、日尺度的森林蒸散发证实了涡度相关法在蒸散量计算上的效果。Denager T 等^[28]分析了丹麦西部野外观测站六年数据,发现入射有效能量平均超过湍流通量,潜热和其他能量通量分量造成了涡动相关法能量不平衡,低估了实际蒸散量。

波文比能量平衡法在假设潜热湍流交换系数和感热湍流交换系数相等的条件下,通过测量实际水汽压的梯度来确定波文比,并应用于能量平衡方程中求解蒸散^[29]。刘凤山等^[30]采用波文比系统研究巨菌草生育盛期的蒸散规律,观测了蒸散量小时变化过程,分析了晴天和雨天的蒸散差异,成功的得出了菌草 ET 气象影响因素。余昭君等^[31]采用单作物系数法和双作物系数法对波文比-能量平衡法进行拟合,结果表明波文比-能量平衡法在半湿润地区葡萄园蒸发蒸腾估算中适用性较高,但是由于半湿润地区近地层垂直温度梯度较大,系统的测定值误差也较高。

蒸渗仪基于水量平衡原理,通过测定容器内部作物和土壤的重量变化来计算蒸散量,在典型作物的小面积蒸散测量和土壤蒸发研究中仍然是最佳选择。Kumar 等^[32]对豌豆进行了田间试验,比较了实测的豌豆需水量和 6 种参考作物蒸散发模型的精度差异,发现蒸渗仪与标准模型的拟合度最高,验证了蒸渗仪在亚温带、亚湿润地区生态水文作物蒸散模型中的应用。刘笑吟^[33]利用小型蒸渗仪开展试验,研究了不同时间尺度和生育期下水稻的蒸散规律,明确了影响稻田蒸散的主要因素和影响程度,文中指出了夜间气温降低导致水汽凝结影响了仪器测量。

遥感法通过遥感技术从气象卫星资料中提取有关生物的参数信息,如植被类型、密度,地表反射的温湿度等。可以对地表温、湿度进行分析反演蒸发,也可以通过植被指数反演作物系数,但是高空的云层和沙尘遮挡会限制遥感的准确性^[34-36]。Huang D 等^[37]使用了伊朗的 Shahrekord 平原、Chaharmahal 省和 Bakhtiari 省主要作物的实际蒸散量与遥感数据结果进行对比分析,结果显示遥感数据计算的蒸散值和实际值吻合度较好。乔雪梅等^[38]使用黑河流域的遥感、气象等数据,获取了当地植被覆盖面积、水域面积和流域湿度指数的变化特征和区域分布并明确了生态环境变化原因。

仪器测量对蒸散发的计算最为直接,但是仪器成本高昂,后期维护也需要人员成本。设备的安装位置和采集环境都有严格的要求,对大区域和长时间单位的蒸散规律缺乏适用性。研究者开始量化各种作物生理、环境参数对蒸散发的内部联系,尝试以数学模型的方式来简化蒸散数据的获取。

1.2.1.2 公式计算

对经验公式的研究源自于 1802 年道尔顿蒸发定律,随后基于公式的蒸散发计算方法迎来了研究热潮。按照计算原理和计算参数可以归纳为基于温度的参考作物蒸散量模型、基于辐射的参考作物蒸散量模型和基于综合法的参考作物蒸散量模型。

基于温度的参考作物蒸散量模型典型代表为 Hargreaves-Samani 模型。Hargreaves-Samani 方法的准确度已经在全球的干旱和半干旱地区的参考作物蒸散量计算中得到了证实,但是在受湿度和风速影响较大地区效果不理想^[39-41]。因此研究者对 HS 参数进行修正。为了探讨在高山地区应用 Hargreaves-Samani 方程计算日参考蒸散量的可能性,Ravazzani G 等^[42]将前人提出的修正方程应用于河流日参考蒸散量,发现计算结果误差较大,在使用局部高程对 Hargreaves-Samani 系数进行了率定后,修正后的方程精准度显著提升,适用于仅使用气温估计高山河流的蒸散估算。刘小飞^[43]根据新乡地区气象要素校正了 H-S 参数,并使用了多年气温数据计算了冬小麦生育期的参考作物蒸散量,拟合结果显示了修正 H-S 公式能较准确的计算华北地区计算参考蒸散量。Makkink、Priestley-Taylor 是基于辐射的参考作物蒸散量模型代表。Chowdhury S 等^[44]采用自动气象站数据计算了 6 个经验模型的蒸散发,在缺乏湿度数据的情况下推荐 Makkink 方法进行蒸散发计算。也有研究者使用 Makkink、Priestley-Taylor 和 Hargreaves 三种方法估算摩洛哥半干旱区参考蒸散发,研究发现原始参数模型已经取得了和实测值一致的趋势,对 Priestley-Taylor 和 Makkink 方程的系数调整后,精度都有不同程度的提升^[45]。

影响蒸散发的因素复杂多样,将空气动力学和能量平衡相组合提出了以 Penman-Monteith 为代表的综合法,P-M 公式综合了外部因素和作物生理特性估算蒸散量,根据空气动力学阻力系数和冠层阻力系数确定作物蒸散量,又经过气象因素的改良,被联合国粮食与农业组织确定为估算参考作物蒸散量的标准方法^[46-49]。国内外学者都对 P-M 公式进行了多地区、多情景的验证,也考虑到几乎不存在风速的温室情况^[50]。刘晓英^[51]将辐射法、气温法、综合法计算的结果与实测值对比,发现吻合程度排名为综合法、辐射、温度法,明确了蒸散影响因子和各模型的内部参数是造成方法差异的原因。闫浩芳等^[52]通过温室微气象数据、黄瓜生理指标等数据对 P-M 模型的冠层组合和空气动力学阻力修正,实现了经验公式在温室环境下的准确应用。

随着蒸散量估算模型的进步和发展,机器学习预测开始进入蒸散发预测领域, Penman-Monteith 公式为国内外蒸散发预测研究提供了重要的验证作用。大量学者针对

不同地理位置、气候条件对公式进行修正计算、类比分析,研究蒸散量的计算精度,建立了相对精确的计算方法,为数值验证提供了理论依据^[53-55]。

1.2.2 机器学习算法在作物需水量预测中的应用

彭曼公式计算过程复杂,所用到的传感器数量多,部分地区气象信息常年缺失、数据难以获取。机器学习算法与传统统计方法相比,只需要较少的假设数量,就能对数据之间的非线性关系建模,并以较高的精度估计非线性函数,许多研究人员利用智能算法建立了基于环境参数的蒸散量预测模型^[56-58]。

机器学习算法划分数据集并使用历史数据进行学习和预测。Hashemi 和 Sepaskhah 评估了 ANN 模型和 P-M 模型在小麦蒸散发估算中的适用性,研究表明,在精确度方面,ANN 的表现优于 P-M 模型^[59]。邢立文等^[60]划分华北地区 60 年时间序列蒸散量数据为训练集和验证集,使用 LSTM 模型模拟成功预测了华北地区的参考作物蒸散量,发现模型在逐月尺度上的模拟要大于日模拟的精度,在内陆地区的模拟要大于沿海地区的模拟精度。为了应对气象数据有限的情况,部分学者将研究重点放在了输入组合对模型的效果上。张薇^[61]训练了 GBDT 预测模型并以 P-M 公式计算结果为基准,发现在最高气温、最低气温和辐射的输入参数组合下模型预测结果与 P-M 计算结果最接近。Yamaç S.S 等^[62]使用气象因素构建输入组合,探讨了 KNN 模型和 ANN 模型在土耳其半干旱环境中效果,结果表明,随着气象因素的增加模拟性能逐步提升,在六个因素组合中的预测精度最高,与 ANN 模型相比,KNN 模型在所有的组合中显示出更好的预测精度和稳定性。Agrawal Y 等^[63]对气象数据进行随机组合,评估了 5 种基于树的机器学习模型(决策树、随机森林、自适应提升、梯度提升机和极端梯度树)的预测性能,结果表明,在所有的月份和高峰月份中极端梯度提升树预测的参考作物蒸散量的最大偏差都小于 0.5 mm/d。考虑到样本平衡问题,防止偏差和过度拟合,许多学者还使用 K 折交叉验证方法训练子集,用多次训练的平均结果评估模型性能^[64,65]。

研究的另一个方面涉及使用组合模型进行训练学习,样本集和模型参数是预测模型的核心,单一模型的超参数主要是根据经验进行设置,对于一些复杂问题,模型很容易陷入过拟合或欠拟合,算法优化后的混合模型显著提高了泛化能力^[66,67]。Tao H 等^[68]比较了萤火虫算法 (FA) 调整的 ANFIS (FA-ANFIS) 模型和经典 ANFIS 模型的预测精度,发现 FA-ANFIS 比经典 ANFIS 模型表现更好。Petković D 等^[69]使用反向传播算法和粒子群算法优化径向基函数网络,在蒸散量预测中,与输入最高气温、最低气温、日照时数、实际蒸汽压和风速的单一模型相比,采用粒子群算法的径向基函数网络具有更好的统计特性。Fan J^[70]对中国不同气候条件的气象数据进行模拟,评估了遗传算法、蚁群优化、布谷鸟搜索算法和花粉授粉算法优化极限学习机在日蒸散量预测方面的性能,研

究发现混合算法都能很好的拟合基准数据，而不同的生物启发式算法之间也存在性能差异。

1.2.3 国内外研究现状总结

国内外学者在需水量估算领域进行了大量不同角度和不同方式的研究，为本文膜下滴灌棉花需水预测的研究提供了重要理论支持和方法指导，现对国内外已有的研究成果进行总结。

(1) 应用蒸散量预测的模型较多的是以 P-M 公式计算值为标准进行对比验证，公式计算的参考蒸散量和实测值尚有部分差距，对实际蒸散量进行模拟的较少，还缺乏不同时间尺度下的预测模型性能研究。

(2) 以往的研究中主要对影响蒸散量的环境因素进行随机组合，研究不同组合之间的差异，环境因素也多以气象数据为主，而作物需水量的影响因素复杂，还需要进一步研究不同因素之间的综合影响。

(3) 机器学习算法应用于不同作物需水量预测的研究较多，但是针对膜下滴灌棉花蒸散量预测的尚不多见，膜下滴灌一定程度上改变了蒸散情况，可以对此进行研究。

1.3 研究目标及研究内容

1.3.1 研究目标

以蒸渗仪实测数据为依据，采用模型仿真的方法，选取斯皮尔曼相关系数(Spearman)和最大互信息系数(MIC)对影响膜下滴灌棉花的多源环境因素进行相关性分析，多角度探索在不同时间尺度下各环境因素对膜下滴灌棉花蒸散量的影响程度，同时通过基于鲸鱼优化算法(WOA)改进的极端梯度提升(XGBoost)模型进行模拟评价，综合计算代价和运行效率获取最优输入特征，并验证模型的可行性。在日尺度下分析预测模型的适应性，并运用多种机器学习算法和群体智能算法对预测模型进行超参数优化，对预测结果进行量化比较，以找到一种契合度高、性能优良的膜下滴灌棉花需水预测模型，为膜下滴灌棉花灌溉决策提供科学依据。结合 Python 算法和 Web 页面设计开发数据管理和模型可视化平台。

1.3.2 研究内容

通过分析调研国内外机器学习算法预测作物需水量的研究现状，本文以蒸渗仪采集的膜下滴灌棉花蒸散量数据为依据，分析环境因素及需水量的关系，探索一种能应用于膜下滴灌棉花需水量的研究中的机器学习模型，通过仿真试验以及误差分析，明确预测

模型最优输入特征和模型参数结构,开展可行性、适用性和有效性对比试验,以期为膜下滴灌棉花灌溉提供参考。并搭建需水管理系统可视化需水信息,具体研究内容如下:

(1) 基于多源环境数据的棉花需水量影响因素分析

影响膜下滴灌棉花需水的因素具有复杂性、非线性的特点,特征维度的大小直接影响了模型计算效率,采用经验选取特征具有随机性,会导致模型的预测精度下降。因此文中采用数据挖掘技术对气象站及传感器采集的环境数据进行数据清洗,采用相关性分析法对环境数据进行不同时间尺度的特征分析,探索膜下滴灌棉花的需水规律和影响因素,以重要性排序结果对环境因素进行组合,分析预测模型在不同输入组合下对棉花需水量的仿真结果,确定棉花混合需水预测模型的最佳输入特征,实现在保证预测精度的基础上,降低数据输入维度,减小预测模型计算成本。

(2) 膜下滴灌棉花需水预测模型建立及优选

针对人工经验调参造成的模型效果偏差问题,基于 XGBoost 模型构建混合膜下滴灌棉花需水预测模型,融合群体智能优化算法优化预测模型超参数,结合特征筛选结果,在不同时间尺度上进行适应度评价,并与 XGBoost、SVM、WOA-SVM、PSO-XGBoost 等模型进行对比,通过仿真模拟进行性能研究,以平均绝对误差(MAE)、决定系数(R^2)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)等为主要评价指标,验证拟合情况,进而对棉花需水预测模型进行评价。

(3) 需水预测模型可视化平台设计与实现

从膜下滴灌棉花灌溉需求出发,基于多源环境数据设计数据库及交互界面,将需水预测模型与系统相结合开发需水预测模型管理系统。实现用户管理、棉花蒸散量预测管理、可视化界面功能,为棉花灌溉决策提供信息化管理。

1.4 技术路线

根据上述研究内容描述,以膜下滴灌棉花需水理论为出发点,采集膜下滴灌棉花数据,对数据进行规约、预处理和划分。结合理论分析和生产实践,对不同尺度的样本集进行相关性分析,探索影响膜下滴灌棉花需水量的影响因素。吸收相关领域研究成果和方法,构建需水预测模型,并以实测值数据进行验证,结合灌溉信息化管理平台,开展膜下滴灌棉花灌溉决策优化研究。本研究总体技术路线如图 1-1 所示。