



华南农业大学学报  
*Journal of South China Agricultural University*  
ISSN 1001-411X, CN 44-1110/S

## 《华南农业大学学报》网络首发论文

题目: 基于物联网和 LSTM 的柑橘园土壤含水量和电导率预测模型  
作者: 高鹏, 谢家兴, 孙道宗, 陈文彬, 杨明欣, 周平, 王卫星  
收稿日期: 2020-07-17  
网络首发日期: 2020-10-19  
引用格式: 高鹏, 谢家兴, 孙道宗, 陈文彬, 杨明欣, 周平, 王卫星. 基于物联网和 LSTM 的柑橘园土壤含水量和电导率预测模型. 华南农业大学学报.  
<https://kns.cnki.net/kcms/detail/44.1110.S.20201016.1702.002.html>



**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。

# 基于物联网和 LSTM 的柑橘园土壤含水量和电导率预测模型<sup>1</sup>

高鹏<sup>1</sup>, 谢家兴<sup>1,2</sup>, 孙道宗<sup>1,2</sup>, 陈文彬<sup>1</sup>, 杨明欣<sup>1</sup>, 周平<sup>1</sup>, 王卫星<sup>1,3\*</sup>

(1. 华南农业大学电子工程学院, 广东广州 510642; 2. 广东省农情信息监测工程技术研究中心, 广东 广州 510642; 3. 广东省智慧果园科技创新中心, 广东 广州 510642)

**摘要:**【目的】构建柑橘果园环境信息物联网实时采集系统, 为果园灌溉施肥管理、效果预测评估提供参考依据。【方法】利用土壤温度、含水量、电导率三合一传感器, 在柑橘果园中设置 5 个节点和 1 个气象站, 通过 ZigBee 短距离无线通信和 GPRS 远距离无线传输, 将果园气象数据和土壤墒情数据传输至远程服务器。利用长短期记忆 (LSTM) 模型建立气象数据与土壤含水量和土壤电导率的预测模型, 计算均方根误差 (RMSE) 和决定系数 ( $R^2$ ) 进行性能评估。【结果】物联网系统能够实现远程传输柑橘果园环境数据, 建立了基于 LSTM 和广义回归神经网络 (GRNN) 的土壤含水量和电导率预测模型, 模型在 5 个节点的数据集的训练结果分别为: LSTM 模型训练的土壤含水量和土壤电导率的 RMSE 范围分别为 6.74~8.65 和 6.68~8.50, GRNN 模型训练的土壤含水量和土壤电导率的 RMSE 范围分别为 7.01~14.70 和 7.60~13.70。利用生成的 LSTM 模型和气象数据进行拟合, 将土壤含水量和土壤电导率的预测值与实测值进行回归分析, LSTM 模型拟合的土壤含水量和土壤电导率  $R^2$  范围分别为 0.760~0.906 和 0.648~0.850, GRNN 模型拟合的土壤含水量和土壤电导率  $R^2$  范围分别为 0.126~0.369 和 0.132~0.268, 说明 LSTM 模型的性能表现较好。【结论】建立了柑橘果园环境的物联网信息传输系统, 构建的基于 LSTM 的果园土壤含水量和土壤电导率预测模型具有较高的精度, 可用于指导柑橘果园灌溉施肥管理。

<sup>1</sup>收稿日期: 2020-07-17; 网络出版时间:

作者简介: 高鹏 (1992—), 男, 博士研究生, E-mail: gaopeng.peng@stu.scau.edu.cn

通信作者: 王卫星 (1963—), 男, 教授, 博士, Email: weixing@scau.edu.cn

基金项目: 广东省重点领域研发计划 (2019B020214003); 广东省科技专项资金 (“大专项+任务清单”) 项目 (2020020103); 广东省教育厅特色创新类项目 (2019KTSX013); 国家荔枝龙眼产业技术体系建设专项资金项目 (CARS-32-14); 广东省现代农业产业技术体系创新团队建设专项资金 (2019KJ108); 广东大学生科技创新培育专项资金项目 (PDJH2019B0080); 大学生创新创业训练计划 (国家级) 项目 (202010564049)

**关键词:** 物联网; 柑橘; 土壤含水量; 土壤电导率; 预测模型

中图分类号: S127 文献标识码: A

## Prediction of soil moisture and soil electrical conductivity based on IoT system and LSTM model

GAO Peng<sup>1</sup>, XIE Jiaying<sup>1,2</sup>, SUN Daozong<sup>1,2</sup>, CHEN Wenbin<sup>1</sup>, YANG Mingxin<sup>1</sup>,  
ZHOU Ping<sup>1</sup>, WANG Weixing<sup>1,3\*1</sup>

(1.College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China; 2.Guangdong Agricultural Information Monitoring Engineering Technology Research Center, Guangzhou 510642, China; 3.Guangdong Intelligent Orchard Science and Technology Innovation Center, Guangzhou 510642, China)

**【Method】** Soil temperature, soil moisture and soil electrical conductivity sensors were applied in five IoT nodes and a weather station was set in citrus orchards. The meteorological data and soil moisture data collected in the orchards were transmitted to a remote server via ZigBee, a short range wireless communication technique, and GPRS, a long distance wireless transmission technique. The prediction models of soil moisture content and soil electrical conductivity were established using weather data based on the long short-term memory (LSTM) model. The root mean square error (RMSE) and coefficient of determination ( $R^2$ ) were calculated to evaluate the performance of the model. **【Result】** The IoT system was capable to transmit environmental data of the citrus orchard to a remote server. LSTM and general regression neural network (GRNN) model were built to predict soil moisture content and soil electrical conductivity. The performance of models in five nodes were as following: The RMSE of soil moisture content and soil electrical conductivity ranged from 6.74 to 8.65 and from 6.68 to 8.50 respectively based on LSTM model, and ranged from 7.01 to 14.70 and from 7.6 to 13.7 respectively based on GRNN model. With the generated LSTM model and meteorological data for predicting, regression analysis was conducted between predicted soil moisture content and soil electrical conductivity and measured values respectively. The  $R^2$  of soil moisture content and soil electrical conductivity ranged from 0.760 to 0.906 and from 0.648 to 0.850 respectively based on LSTM model, and ranged from 0.126 to 0.369 and 0.132 to 0.268 respectively based on GRNN model. The results indicated that the

LSTM model performed better than the GRNN model. 【Conclusion】 The IoT system for citrus orchard environmental information transmission is established. The LSTM model has high accuracy in predicting soil moisture content and soil electrical conductivity, and the model can be helpful for guiding irrigation and fertilization management.

**Key words:** internet of things; citrus; soil moisture; soil electrical conductivity; prediction model

柑橘 *Citrus*spp. 是一种重要的经济作物，主要分布在我国的广东、广西和四川等地<sup>[1-3]</sup>。我国的柑橘种植面积和产量已经位居世界首位，并且连续多年不断增长<sup>[4,5]</sup>。目前我国的柑橘灌溉由传统的漫灌逐渐升级到滴灌，能够较好的实现节水灌溉<sup>[6]</sup>。但是滴灌作业主要还是由人为主观意识根据气象状况来进行判断，缺乏客观依据<sup>[7,8]</sup>。柑橘生长需要适宜的土壤含水量和电导率，过低或者过高的土壤含水量都会对根系生理活动产生负面影响<sup>[9]</sup>，而土壤电导率则直接反映土壤的含盐量程度<sup>[10]</sup>，过高的盐胁迫会导致柑橘产生盐毒害，过低的盐含量导致柑橘难以吸收所需的无机盐离子。因此，在柑橘种植作业中，建立能够及时获取柑橘果园的气象数据、土壤数据，建立果园的土壤含水量和土壤电导率预测模型，对于制定和指导果园灌溉施肥管理具有重要意义。

近年来，物联网技术在农业中获得了广泛的应用，结合相关传感器能够实现果园生长环境的远程实时监测。杨伟志等<sup>[11]</sup>建立了基于物联网和人工智能的柑橘灌溉系统，该研究利用传感器实时监测柑橘果园土壤水分等信息，在物联网系统传输下，与专家系统知识库的灌溉规则进行综合来制定灌溉控制措施，该系统主要依靠已有的控制规则，部分参数的设置需要依靠经验值来确定。李中良等<sup>[12]</sup>利用物联网和空气温湿度传感器、土壤温湿度传感器设计了实时监测柑橘环境的系统，该系统综合了无线传感器网络、通用无线分组服务（General packet radio service, GPRS）网络与远程服务器网络来实现田间数据的远程传输，但该系统没有针对果园环境数据进行进一步的研究和分析。罗党等<sup>[13]</sup>利用灰色神经网络，研究降雨量、温度、湿度、风速等因子对土壤含水量的影响，该模型说明了神经网络在土壤含水量预测方面是可行的，具有不错的精度。蔡亮红等<sup>[14]</sup>人设计了基于极限学习机（Extreme learning machine, ELM）算法的土壤含水量预测模型，该研究采用定点采样土壤样本并获取光谱数据的方法，经过小波变换和特征工程处理后，利用 ELM 算法建立非线性模型，实现土壤含水量的预测。但该系统没有利用物联网等实时传输方法，而且加入的训练因子较少，在复杂环境条件下，还需要进一步研究。张武等<sup>[15]</sup>人利用反向传播（Back propagation, BP）神经

网络和长短期记忆（Longshort-term memory，LSTM）深度学习模型研究了不同采样间隔对茶园土壤墒情预测模型性能的影响，结果表明 30min 的采样间隔下，模型的性能表现最好，但是该研究的数据采集时间跨度较短，模型也未给出更多可解释性指标来描述其模型性能。

针对以上研究现状，本文设计一种基于物联网的柑橘园环境信息实时采集与传输系统，并建立基于 LSTM 的土壤含水量和电导率预测模型，为果园灌溉施肥管理提供指导。

1 柑橘园环境信息物联网传输系统设计

1.1 系统总体架构

物联网传输系统<sup>[16]</sup>包括远程服务器、数据库、无线通信模块、信息采集节点、气象站等。信息采集节点负责采集柑橘园中的土壤含水量和电导率信息，气象站负责采集果园中的气象信息，包括空气温度、湿度、风速、降雨量等，通过无线通信模块传输至物联网网关与远程服务器，传输的数据最终存储在数据库中<sup>[17]</sup>。系统架构如图 1 所示。

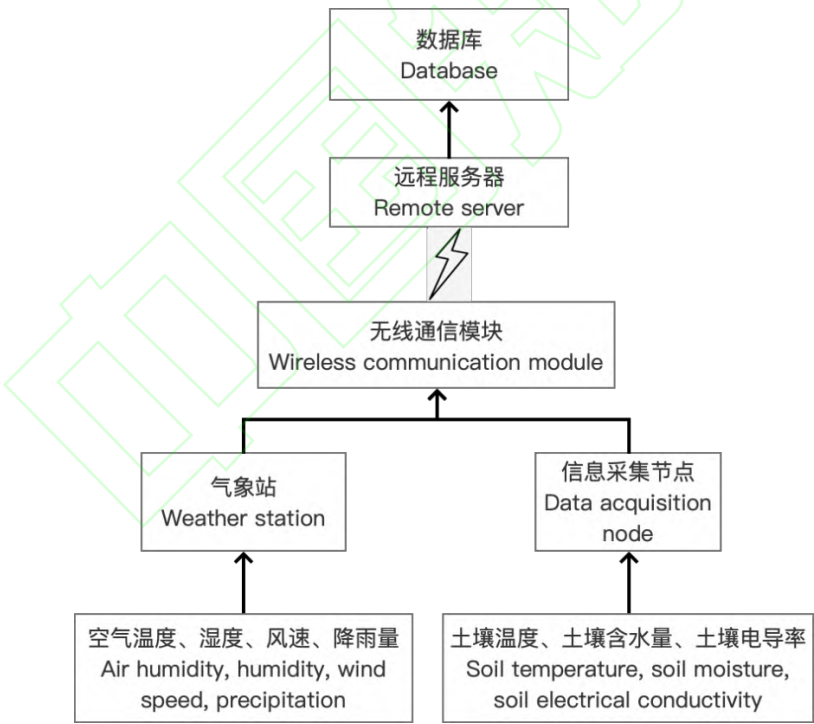


图 1 物联网传输系统架构图

Fig.1 Framework of IoT transmission system

1.2 物联网传输系统硬件

硬件系统包括 STM32F103C8T6 处理器模块、电源模块、GPRS 模块、ZigBee 模块、空气温湿度传感器、风速传感器、降雨量传感器和土壤温度、含水量、电导率三合一传感器。传感器模块用于直接采集相关果园环境信息，ZigBee 模块负责各个节点与网关数据传输，GPRS 模块将各个节点数据远程传输至远程服务器。硬件框图如图 2 所示。

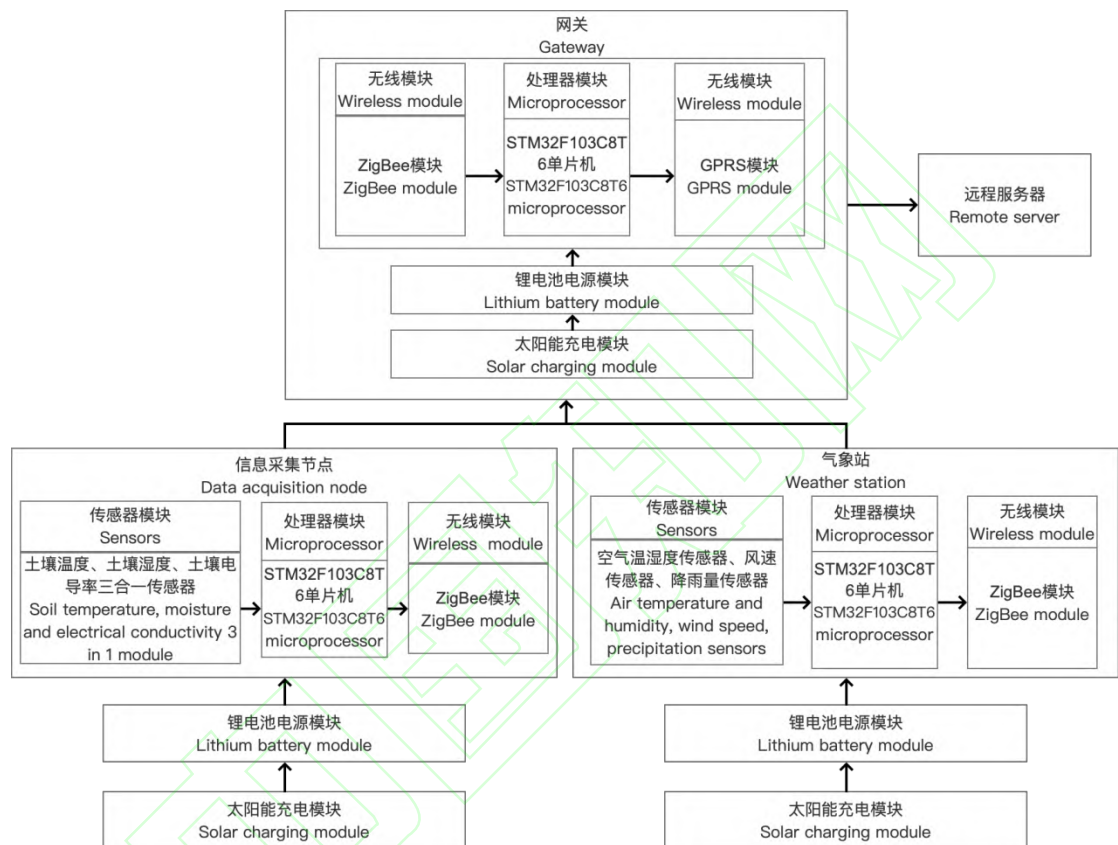


图 2 物联网传输系统硬件框图

Fig.2 Hardware framework of IoT transmission system

### 1.3 物联网传输系统软件

#### 1.3.1 ZigBee 组网协议

常见的 ZigBee 组网协议主要包括树型组网、星型组网和网状型组网<sup>[18,19]</sup>。本文信息采集节点和气象站布置于柑橘果园中，需要考虑节点数据传输性能和续航。由于星型组网方式



下，各个节点需要与网关直连以传输数据，在果园中存在着无法传输的可能。网状型组网方式需要经过多轮数据转发，容易引起额外的能量损耗和丢包。因此本文采用树型组网方式<sup>[20]</sup>，能够利用少量的节点中继实现数据传输，也能够尽可能减少路由转发引起的丢包问题。

1.3.2 节点软件

系统节点主要负责数据采集与转换，其包含两种工作模式，即正常模式和休眠模式<sup>[21,22]</sup>。在正常工作模式下，节点负责开启传感器并采集数据；而在休眠模式下，节点仅保留时钟唤醒机制，定时开启并采集数据，以实现节能的目的。节点软件工作流程图如图 3 所示。

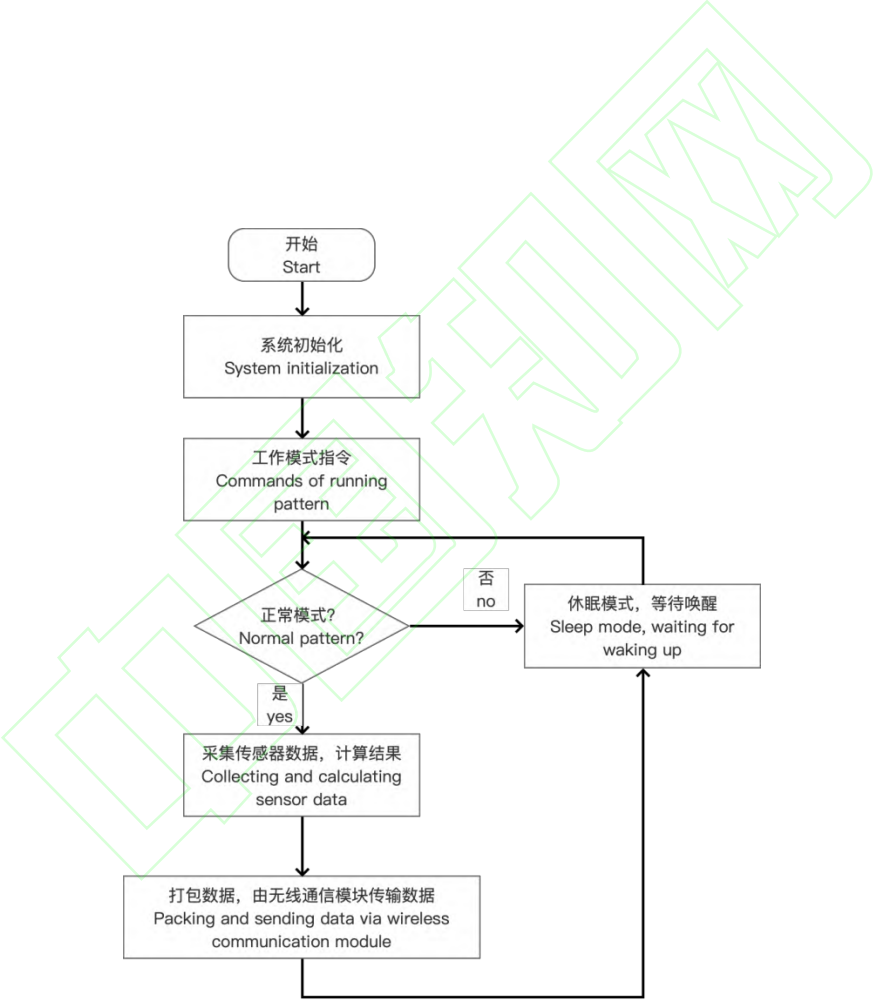


图 3 软件工作流程图

Fig.3 Flowchart of software

2 材料与方法

2.1 区域概况与数据采集预处理

试验地点设在华南农业大学试验基地，该基地面积约 13hm<sup>2</sup>，海拔约为 200m，属于山地地形，主要为含沙土质。该地区夏季高温多雨，年降雨量为 1700mm 左右，但是冬季降雨较少，易干旱。昼夜温差较大，适合柑橘等作物积累糖分。在果园设置了 5 个节点，气象站和网关位于果园中央位置；使用激光测距仪测量得到单节点之间最长距离为 344m，节点使用高度为 2.5m 的支架进行固定以减少果树对无线信号传播的干扰。其布置示意图如图 4 所示。节点数据与气象站数据分别传输至网关，由网关负责将数据打包传输至远程服务器，完成数据的传输。



图 4 节点布置位置图

Fig.4 Figure of node location

本文的系统搭建完成后正式数据采集时间从 2019 年 10 月 1 日至 2020 年 6 月 16 日，节点数据传输间隔设置为每 10 min 传输 1 个数据包，单个节点理论上每天的数据包量为 144 个。数据采集期间平均每个节点接收数据包总数为 36326 个，总平均丢包率为 2.6%。利用汇总后的数据包统计当天的最高温度、最低温度、平均温度、湿度、降雨量、土壤含水量、土壤电导率和土壤温度。土壤含水量和土壤电导率一般与蒸散量、降雨量、温度、湿度等相关，根据气象参数和彭曼公式计算蒸散量<sup>[23]</sup>如下：

$$ET_0 = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)} \quad (1)$$



式中， $ET_0$ 为日参考蒸散量， $\text{mm}\cdot\text{d}^{-1}$ ； $R_n$ 为净辐射量， $\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ ； $G$ 为土壤热通密度， $\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ ； $T$ 为2m高度的平均空气温度， $^{\circ}\text{C}$ ； $u_2$ 为2m高度的风速， $\text{m}\cdot\text{s}^{-1}$ ； $e_s$ 为饱和蒸汽压， $\text{kPa}$ ； $e_a$ 为实际蒸汽压， $\text{kPa}$ ； $\Delta$ 为蒸汽压力曲线斜率， $\text{kPa}\cdot^{\circ}\text{C}^{-1}$ ； $\gamma$ 为温湿度常数， $\text{kPa}\cdot^{\circ}\text{C}^{-1}$ 。采用估算公式（2）计算辐射数据<sup>[24]</sup>：

$$R_s = \left(a_s + b_s \frac{t}{t_{\max}}\right) R_a \quad (2)$$

式中， $R_s$ 为太阳辐射， $\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ ； $t$ 为日照时长， $\text{h}$ ； $t_{\max}$ 为最大日照时长， $\text{h}$ ； $a_s$ 和 $b_s$ 通常取常数0.25和0.50。

由于各个环境参数具有不同的数量级，因此需要在数据预处理环节进行归一化处理。各个节点预处理后的数据取平均值后，分别建立各个环境参数与土壤含水量和土壤电导率的相关性分析<sup>[25]</sup>，结果如表1所示。

表 1 土壤含水量和电导率与环境参数的相关系数

Table 1 Correlation analysis of soil moisture and soil electrical conductivity

| 土壤因子           | 最高温<br>度               | 最低温<br>度               | 平均温<br>度                | 降雨量               | 空气<br>湿度            | 土壤温<br>度                | 蒸散量                    |
|----------------|------------------------|------------------------|-------------------------|-------------------|---------------------|-------------------------|------------------------|
| Soil<br>factor | Max<br>temperat<br>ure | Min<br>temperat<br>ure | Mean<br>temperat<br>ure | Precipitat<br>ion | Air<br>humidi<br>ty | Soil<br>temperat<br>ure | Evapotranspira<br>tion |
| 含水量            | 0.490                  | 0.598                  | 0.644                   | 0.123             | 0.430               | 0.652                   | 0.158                  |
| moisture       |                        |                        |                         |                   |                     |                         |                        |
| 电导率            | 0.467                  | 0.311                  | 0.231                   | 0.300             | 0.168               | 0.251                   | 0.148                  |
| Electrical     |                        |                        |                         |                   |                     |                         |                        |

conductiv

ity

由表 1 可以发现，气象环境因子与土壤含水量和土壤电导率存在一定的相关性，具备输入模型参与拟合的条件。

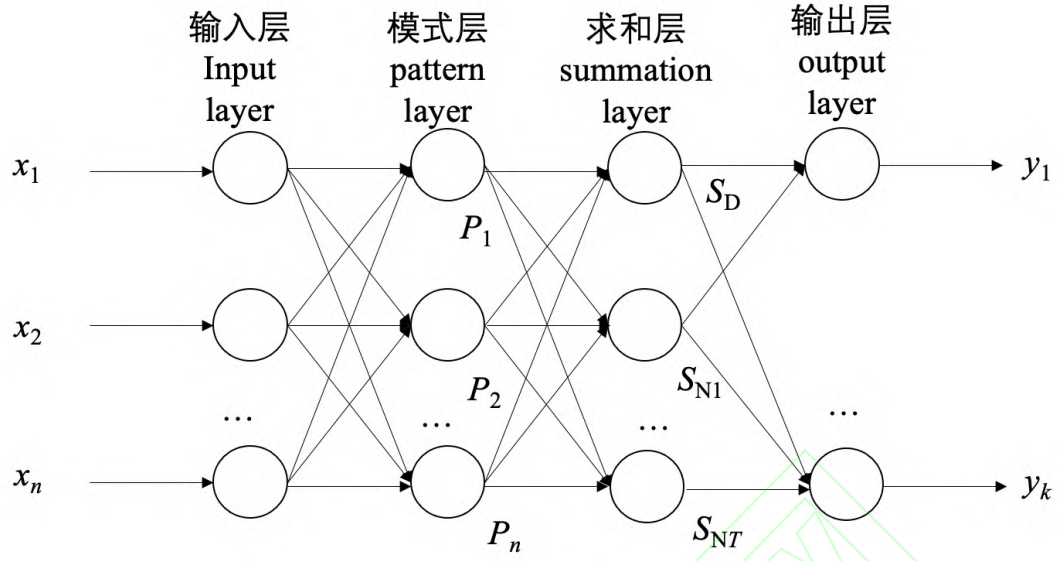
## 2.2 GRNN 模型

GRNN 模型最早是由 Specht<sup>[26]</sup>提出的。GRNN 模型能够利用连续的非线性数据进行训练，并收敛于底层回归曲面，该算法能够应用在学习植物模型的动力学中。GRNN 建立在非参数回归基础上，以样本数据为后验条件，执行 Parzen 非参数估计，依据最大概率原则计算网络的输出并传输给下一级。GRNN 的基础是径向基网络，具有较好的非线性拟合功能，但在训练过程中需要选定合适的学习参数，否则会容易产生过拟合现象。GRNN 的网络结构分为 4 层，分别是输入层，模式层，求和层和输出层。其基础结构图如图 5 所示。

GRNN 的网络输入为矩阵  $X=[x_1, x_2, \dots, x_n]^T$ ，输出为矩阵  $Y=[y_1, y_2, \dots, y_k]^T$ ， $k$  为输出向量的维数。GRNN 的模式层的神经元数目等于输入样本数目  $n$ ，该层的传输函数为公式 (3)：

$$p_i = \exp \left[ -\frac{(X-X_i)^T(X-X_i)}{2\delta^2} \right] \quad i = 1, 2, \dots, n \quad (3)$$

式中，神经元  $i$  的输出为输入变量与其对应的样本  $X$  之间欧几里得距离平方的指数平方  $(X-X_i)^T(X-X_i)$  的指数形式。式中， $X$  为网络输入变量， $X_i$  为第  $i$  个神经元对应的学习样本， $\delta$  为平滑参数，改变  $\delta$  即可改变网络学习率。



$x_1 \sim x_n$ : 输入样本,  $P_1 \sim P_n$ : 各输入样本对应的传递函数值,  $S_D$ : 算术求和,  $S_{N1} \sim S_{NT}$ : 加权求和,  $y_1 \sim y_k$ : 各个维度的输出样本

$x_1 \sim x_n$ : Input samples,  $P_1 \sim P_n$ : The transfer function values of samples,  $S_D$ : Arithmetic summation,  $S_{N1} \sim S_{NT}$ : Weighted summation,  $y_1 \sim y_k$ : The output samples of each dimension

图 5 GRNN 网络结构

Fig.5 Structure of GRNN model

GRNN 网络求和层包含 2 种求和方式, 即直接求和与加权求和。2 种方式的公式分别如公式 (4) 和 (5) 所示:

$$S_D = \sum_{i=1}^n \exp \left[ -\frac{(x-x_i)^T(x-x_i)}{2\sigma^2} \right], \quad (4)$$

$$S_{Nj} = \sum_{i=1}^n y_{ij} P_i, \quad j = 1, 2, \dots, k, \quad (5)$$

式中,  $S_D$  为求和层神经元输出,  $S_{Nj}$  为求和层加权神经元输出,  $i$  表示模式层中第  $i$  个神经元,  $j$  表示输出样本  $Y$  中第  $j$  个元素,  $P_i$  为模式层第  $i$  个神经元的权值。

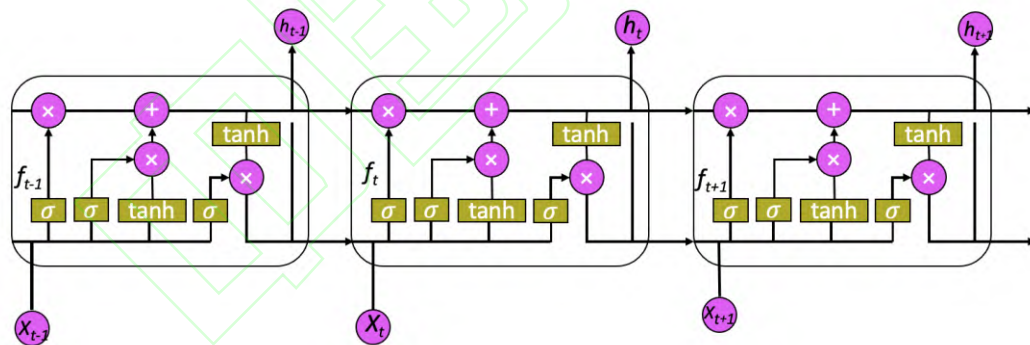
GRNN 网络的输出层中神经元数目等于输出向量的维数, 如公式 (6) 所示。

$$y_j = \frac{S_{Nj}}{S_D}, \quad j = 1, 2, \dots, k. \quad (6)$$

## 2.3 LSTM 模型

实践证明，神经网络能够有效地解决复杂的非线性问题，但传统的神经网络无法利用时间序列数据特征预测未来的情况，特别是对市场销售、股票价格、pm2.5 等时间序列数据的时序建模。时间序列数据不仅与当前条件相关，而且与最后的数据具有很高的相关性，而传统的神经网络难以对时间序列进行建模。近年来，递归神经网络 (Recursive neural network<sup>[27-29]</sup>, RNN) 在自然语言处理、机器翻译和文本预测等方面得到了成功的应用，其效果明显优于神经网络。RNN 不同于一般的前馈网络，它包含隐藏状态提取特征，然后转化为输出数据。另一方面，RNN 在所有计算步骤中共享权值、偏差等，使其与所连接的节点相依赖。然而，标准 RNN 也有其不足之处，如要求输入和输出数据大小相同。另外，通过对前一层隐状态的变换和激活函数得到 RNN 每一层的隐层状态，使其容易发生，即，当 RNNs 对反向传播的某些阶导数时的梯度消失和梯度爆炸问题。因此，RNN 有时不能很好地处理长期依赖问题。

作为一种特殊的 RNN 网络，LSTM<sup>[30]</sup> 的区别在于其包含了一个控制应该丢弃哪些信息的门，即遗忘门 (forget gate)。其次，为了避免梯度消失问题，在计算隐层状态时，用求和代替梯度导数。因此，LSTM 具有解决长期问题的理论能力。在本文中，LSTM 网络是由 3 个部分组成的，即一个输入层用于处理数据，50 个隐藏层用于训练和建模，一个输出层用于输出建模结果。模型的核心部分即 LSTM<sup>[31]</sup> 节点，主要由 3 个部分组成，分别是输入门，输出门和遗忘门。其结构如图 6 所示。



$f_t$ : 遗忘门丢弃的信息,  $\sigma$ : Sigmoid 函数,  $\tanh$ : 双曲正切函数,  $h_t$ : LSTM 单元输出内容

$f_t$ : The information got rid of,  $\sigma$ : Sigmoid function,  $\tanh$ : tanh function,  $h_t$ : The output of LSTM cell

图 6 LSTM 结构图

Fig. 6 Structure of LSTM

其中，遗忘门控制丢弃的信息，如公式（7）所示：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (7)$$

式中， $f_t$ 的范围为[0,1]， $\sigma$ 函数使得方程非线性化，更接近实际值从而提高预测精度。 $W_f$ 和 $b_f$ 分别为权重和偏差参数。 $h_{t-1}$ 和 $x_t$ 为上一状态值和当前输入值。

输入门的功能是控制输入到当前单元中的信息，其状态更新如公式（8）和（9）所示：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (8)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (9)$$

式中， $i_t$ 由 $\sigma$ 激活函数控制输入数据，如果其值为0，表示输入数据被完全阻塞，如果值为1，则表示完全允许数据进入门中。细胞状态由 $\tilde{C}_t$ 负责，它的激活函数是 $\tanh$ 。 $b_i$ 和 $b_C$ 表示函数偏差参数。

输入门经过运算的数据，需要更新到细胞状态中，更新公式为：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (10)$$

式中， $f_t * C_{t-1}$ 代表遗忘的信息， $i_t * \tilde{C}_t$ 代表需要更新到当前细胞状态的信息，两者之和共同决定向下一个状态传递的信息量。

最后一部分是输出门，其更新公式为：

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (11)$$

$$h_t = o_t * \tanh(C_t), \quad (12)$$

式中，输出的信息由 $\sigma$ 函数决定，同时更新后的细胞状态值 $C_t$ 在 $\tanh$ 函数作用下，与传递函数 $o_t$ 的乘积作为向下一个单元输出的内容， $b_o$ 表示传递函数偏差参数。

经过上述 LSTM 单元的运算后，输出的信息能够被存储，也能够进入下一个节点进行运算，从而在训练过程中提取出信息。

## 2.4 模型性能测试标准

本文模型的性能表现使用均方根误差 RMSE 来表示，计算模型预测值与真实测量值的决定系数  $R^2$ <sup>[32]</sup>来表征模型可解释性。相关公式为：

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{pred} - y_{cal})^2}{N}}, \quad (13)$$



$$R^2 = \frac{\sum_{i=1}^N (y_{\text{pred}} - \overline{y_{\text{cal}}})^2}{\sum_{i=1}^N (y_{\text{cal}} - \overline{y_{\text{cal}}})^2}, \quad (14)$$

式中,  $y_{\text{pred}}$  表示预测值,  $y_{\text{cal}}$  表示实测值,  $\overline{y_{\text{pred}}}$  表示预测值均值,  $\overline{y_{\text{cal}}}$  表示实测值均值,  $N$  表示样本数。

### 3 结果与分析

#### 3.1 LSTM 和 GRNN 模型训练设置

试验中, 使用了果园气象站气象数据和 5 个信息采集节点采集到的土壤墒情数据, 现场布置的节点如图 7 所示。由于无线传感器网络传输过程中存在复杂因素的干扰, 产生了个别的异常(如空数据)与丢包数据, 使得数据包不完整或整体丢失, 本文将异常数据进行剔除, 最终得到符合模型需求的数据。将 70%数据划分为训练集, 30%数据划分为测试集, 5 个节点数据划分结果如表 2 所示。训练参数设置方面, LSTM 模型包含 2 层, 每层 128 个 LSTM 单元, 使用 Relu 激活函数, 损失函数使用均方误差 (Mean Square Erroe, MSE), 优化算法使用 Adam, 共设置 72 个 epoch 和 128 个 batch size。GRNN 模型包含 4 层, 模式层与输入样本相同, 求和层使用 128 个神经元, 激活函数、损失函数和优化算法、epoch 和 batch size 等参数与 LSTM 模型保持一致。

表 2 节点训练数据信息

Table2 Datasets of nodes

| 节点   | 数据总数         | 训练集              | 测试集             |
|------|--------------|------------------|-----------------|
| Node | Total number | Training dataset | Testing dataset |
| 1    | 216          | 151              | 65              |
| 2    | 241          | 168              | 73              |
| 3    | 241          | 168              | 73              |
| 4    | 196          | 137              | 59              |
| 5    | 203          | 142              | 61              |

## Weather station



图 7 柑橘果园节点安装图

Fig.7 Fig of nodes installation in citrus orchard

## 3.2 模型性能表现

基于本文所提出的模型和预处理的输入样本经过训练后的性能测试标准如表 3 所示。

表 3 模型性能表现

Table3 Performance of models

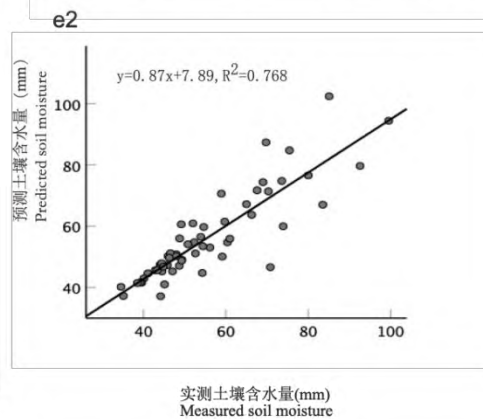
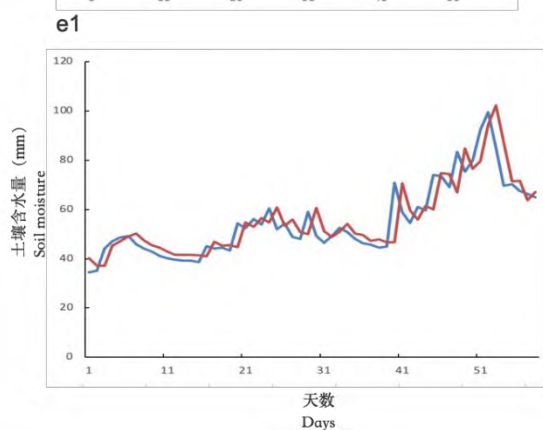
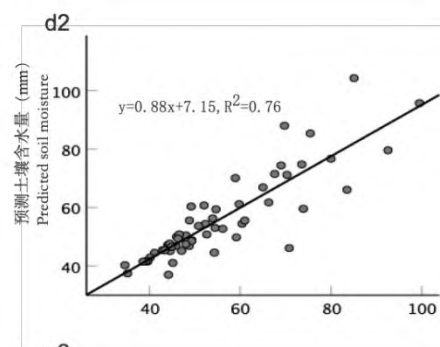
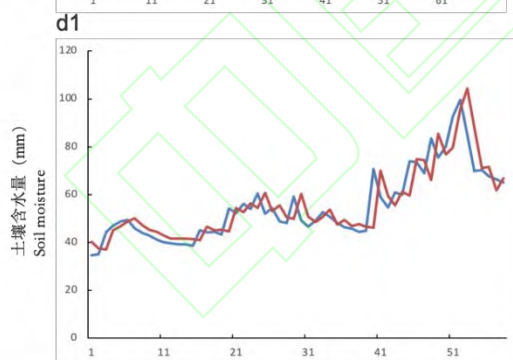
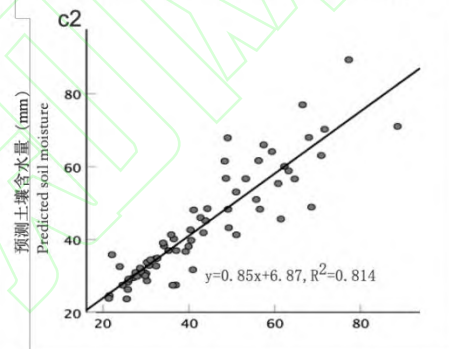
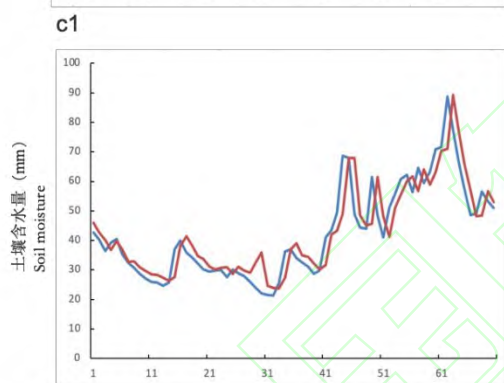
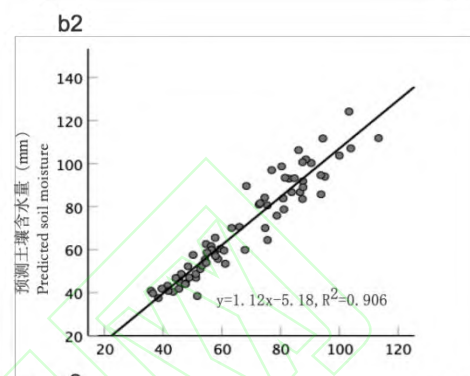
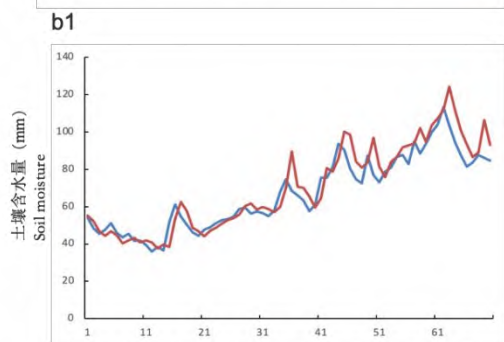
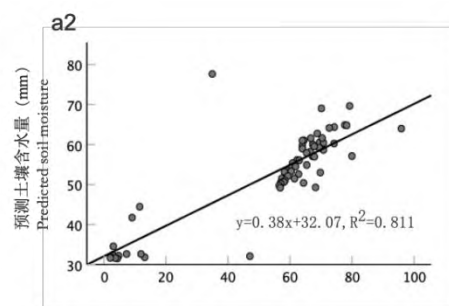
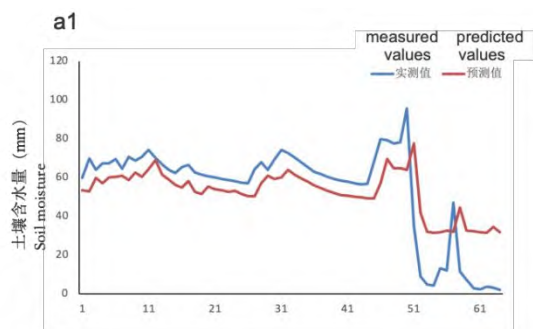
| 节点   | 模型    | 模型表现              |       |                              |       |
|------|-------|-------------------|-------|------------------------------|-------|
| Node | Model | Model performance |       |                              |       |
|      |       | 土壤含水量             |       | 土壤电导率                        |       |
|      |       | Soil moisture     |       | Soil electrical conductivity |       |
|      |       | RMSE              | $R^2$ | RMSE                         | $R^2$ |
| 1    | LSTM  | 8.65              | 0.811 | 6.88                         | 0.85  |
|      | GRNN  | 13.3              | 0.126 | 12.3                         | 0.106 |
| 2    | LSTM  | 7.3               | 0.906 | 6.68                         | 0.727 |
|      | GRNN  | 14.7              | 0.323 | 13.7                         | 0.185 |
| 3    | LSTM  | 6.74              | 0.814 | 8.22                         | 0.648 |
|      | GRNN  | 7.01              | 0.369 | 8.26                         | 0.268 |
| 4    | LSTM  | 7.93              | 0.76  | 8.5                          | 0.767 |
|      | GRNN  | 10.27             | 0.256 | 7.85                         | 0.167 |
| 5    | LSTM  | 8.4               | 0.768 | 7.78                         | 0.769 |
|      | GRNN  | 8.8               | 0.206 | 7.6                          | 0.132 |

从表 3 中可以看出，在土壤含水量的预测模型表现中，LSTM 模型的 RMSE 值范围从 6.74~8.65，低于 GRNN 模型的 7.01~14.7。LSTM 模型中，节点 2 和节点 3 的 RMSE 在 5 个节点中较低，这可能是因为两组训练样本数较多。而节点 4 的 RMSE 比节点 5 更低，但其训练样本数比后者少，因此 RMSE 除了受训练样本数影响外，还与数据集自身样本情况有关。GRNN 模型的 RMSE 值与训练样本数没有呈现明显的关系或趋势。RMSE 用来衡量离散程度，与原样本具有相同的量纲，而决定系数  $R^2$  作为一种无量纲的性能指标，能够反映模型的可解释性。从表 6 可以看出，在土壤含水量模型表现中，虽然 LSTM 模型的 RMSE

相比于 GRNN 模型没有表现出明显优势,但是其  $R^2$  范围从 0.76~0.906,远好于 GRNN 模型的 0.126~0.369,这意味着 LSTM 模型预测的土壤含水量呈现出了较高的可解释性,意味着该模型给出的预测值能够较好的反映土壤含水量。表 6 中土壤电导率模型预测表现与土壤含水量呈现类似的趋势,同样表明 LSTM 模型在给定的试验样本下,适合用来对土壤电导率进行预测。

### 3.3 模型拟合分析

LSTM 模型的预测精度较高,性能稳定,故而使用该模型观察预测效果。预测方法是:利用模型和气象站气象数据,获取模型生成的预测值,与物联网信息节点采集到的实际数据进行回归分析。绘制的对比结果如图 8 和图 9a1~e1 所示,之后利用最小二乘法进行回归拟合,其结果如图 8 和图 9a2~e2 所示。其中,5 个节点分别布置在果园 5 个分区中,由于果园面积较大,同时不同区域有高低落差,因此 5 个节点数据分别代表果园各个分区的土壤墒情。





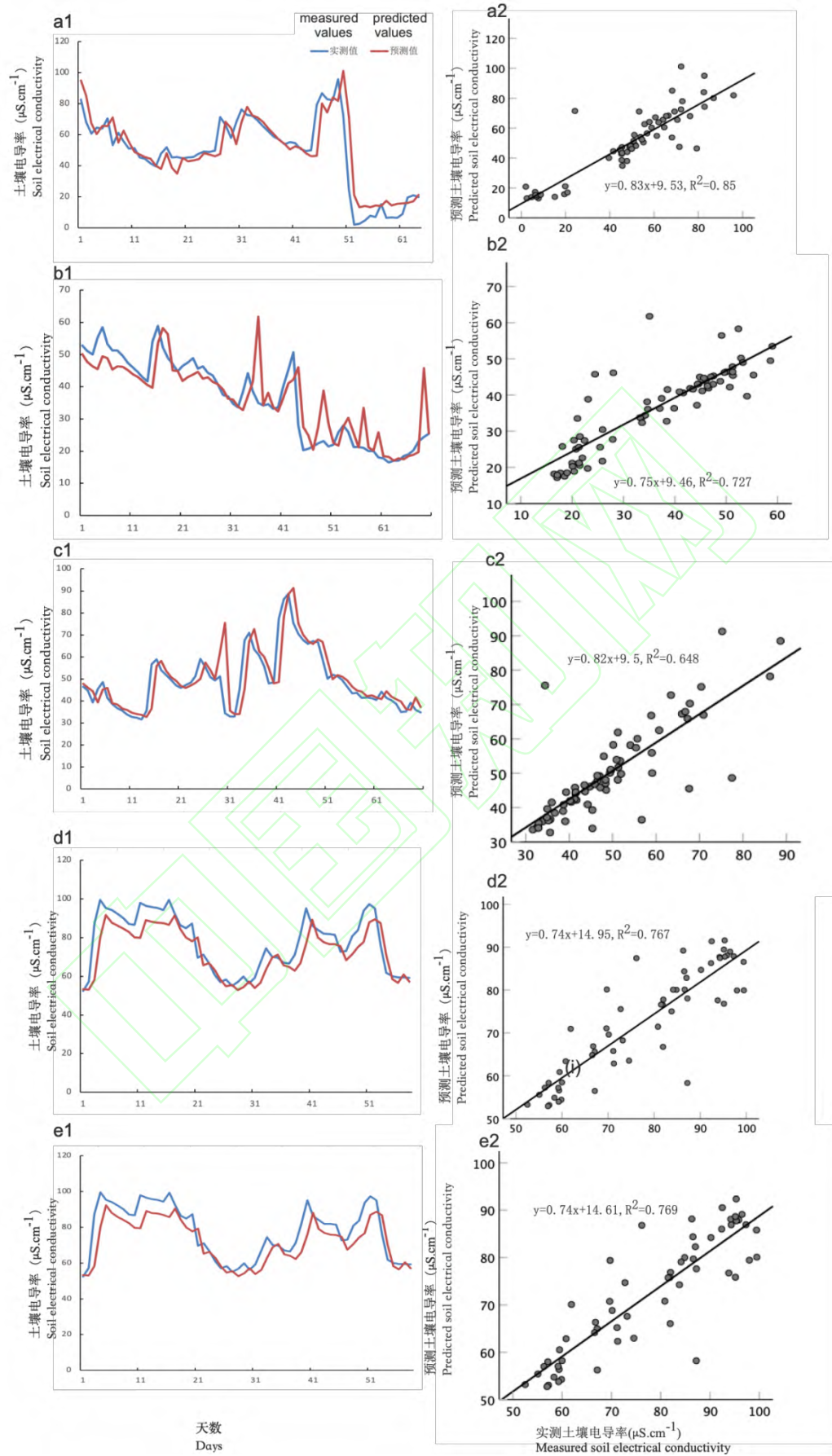
a1~e1 为实测值与预测值对比, a2~e2 为实测值与预测值回归拟合

a1~e1 represent the comparisons of measured and predicted values, a2~e2 represent the regression fittings of measured and predicted values

图 8 LSTM 模型在节点 1~5 的土壤含水量预测值与实测值的对比

Fig.8 Comparison of measured and predicted soil moisture of LSTM model in nodes 1~5





a1~e1 为实测值与预测值对比, a2~e2 为实测值与预测值回归拟合

a1~e1 represent the comparisons of measured and predicted values, a2~e2 represent the regression fittings of measured and predicted values

图 9 LSTM 模型在节点 1~5 的土壤电导率预测值与实测值的对比

Fig.9 Comparison of measured and predicted soil electrical conductivity of LSTM model in nodes 1~5

由图 8a 和 8f 可以看出，预测值相对于实测值偏小，而拟合得到的回归方程中，实测的土壤含水量数据点主要集中在 56~76 mm 区间，这个趋势在图 8a 中的折线图中无法反映出来，这可能是由于节点 1 处在相对低洼的位置，加上外界环境变化如大量降雨积水或者人工灌溉导致的短时间内土壤含水量升高，在图 9a 中可以观察到此时的土壤电导率在土壤含水量升高后也迅速下降的现象。图 8g 的回归拟合决定系数高达 0.906，观察图 8b 中的折线图可以看出，预测值很好地贴近了实测的土壤含水量。图 8c 与图 8h 的拟合效果与节点 2 即图 8b 和图 8g 接近，说明模型在节点 2 和节点 3 的数据集中能够获得较好的训练结果。节点 4 和节点 5 的模型拟合效果如图 8i 和图 8j 所示，其决定系数分别为 0.76 和 0.768，低于 0.8，根据回归方程图可以看到，主要是由于个别数据相对于其他样本偏差较大导致的。这两个节点对应的模型的均方根误差 RMSE 并不是最大的，这也再次说明了仅根据 RMSE 指标不能证明模型的性能高低。

从图 9 中可以看出，节点 3 的土壤电导率预测模型结果即图 9c 和图 9h 的数据点多集中在 30~55  $\mu\text{S}\cdot\text{cm}^{-1}$  区间，从折线图中可以看到预测值线与实测值线比较接近，而其回归拟合决定系数仅为 0.648，说明模型预测性能会受到数据分布的影响，较多的数据量可能会减少此种影响。虽然使用了同样的气象数据，但土壤电导率的预测模型性能总体上稍低于土壤含水量预测模型，这可能是由于土壤电导率需要除了受气象因子、柑橘果树蒸散发影响外，还受到人工施肥、生物无机盐代谢等多种因素的影响，这可能是需要进一步研究的方向。

#### 4 结论

本文提出了利用物联网来实现远程实时获取柑橘果园土壤含水量和土壤电导率的方法，并搭建气象站获取柑橘果园气象信息，利用 LSTM 和 GRNN 模型对土壤墒情进行建模，实现对柑橘园土壤含水量与土壤电导率的预测。部分已有的对土壤墒情进行建模的研究，以

BP 神经网络或光谱反演方法为基础,存在着实时性不足、长时间预测精度较低的问题。本文利用物联网及相关传感器搭建远程获取果园环境数据的系统,在果园中设置 5 个土壤墒情采集节点和 1 个气象站,实现了实时采集和传输果园环境数据与气象数据的目的。

利用采集到的气象数据与节点土壤墒情数据在 LSTM 模型和 GRNN 模型上的训练结果表明: LSTM 模型训练的土壤含水量和土壤电导率的 RMSE 范围分别为 6.74~8.65 和 6.68~8.50,而其决定系数  $R^2$  范围分别为 0.76~0.906 和 0.648~0.850,相比于 GRNN 模型的 RMSE 和  $R^2$  (7.01~14.70 和 7.60~13.70, 0.126~0.369 和 0.132~0.268),表现出了较好的模型性能与预测效果。在 LSTM 模型的土壤含水量和土壤电导率拟合分析中可以发现,模型的 RMSE 与  $R^2$  之间并没有明确的相关关系,即较小的 RMSE 并不意味着较高的决定系数,因此只用 RMSE 来衡量模型性能可能是片面的。

LSTM 模型的预测值与实测值的拟合回归方程和对应的折线图说明模型的预测效果较好,能够满足预测的需求。但是部分节点如节点 3 的土壤电导率拟合决定系数较低,说明数据样本分布可能会对模型效果产生影响。本文建立的 LSTM 预测模型在土壤电导率的预测性能方面稍逊于土壤含水量模型,这可能是由于土壤电导率受到的环境中其他因素影响更多,需要更进一步的深入研究。另外,果园的土壤含水量和土壤电导率在不同的种植和生理条件下,存在一定的差异,根据生产过长中的灌溉与施肥情况,利用本文提出的方法能够为柑橘果园的灌溉和施肥提供一定的指导,也可以根据 LSTM 预测模型,结合未来一段时间气象数据与当前田间土壤墒情数据,预测未来一段时间的土壤墒情数据,及时掌握土壤墒情变化趋势,方便果园管理者提前做好施肥灌溉准备工作,能够帮助实现科学精细的果园灌溉与施肥管理,带来较好的管理效益。

## 参考文献

- [1] 成家壮, 韦小燕, 范怀忠. 广东柑橘疫霉研究[J]. 华南农业大学学报, 2004(2): 31-33.
- [2] 张超博, 李有芳, 李思静, 等. 土壤管理方式对伏早期柑橘生长及土壤温度和水分的影响[J]. 华南农业大学学报, 2019, 40(3): 45-52.
- [3] 肖畅, 彭婷, 刘继红. 基于 WOS 和 CiteSpace 分析我国近十年柑橘研究热点与前沿[J]. 果树学报, : 1-17.
- [4] 李莉萍. 柑橘产业化经营案例分析:评《柑橘绿色生产与产业化经营》[J]. 中国果树, 2020(3): 151.

- [5] 罗省根, 双巧云, 陈团显, 等. 江西省柑橘出口现状及发展对策[J]. 中国果树, 2020(3): 138-140.
- [6] 赖俊桂, 孙道宗, 王卫星, 等. 基于无线传感器网络的山地柑橘园灌溉控制系统设计与试验[J]. 江苏农业科学, 2020, 48(7): 245-249.
- [7] 陈昱辛, 贾悦, 崔宁博, 等. 滴灌水肥一体化对柑橘叶片光合、产量及水分利用效率的影响[J]. 灌溉排水学报, 2018(S2 vo 37): 50-58.
- [8] 谢家兴, 高鹏, 莫昊凡, 等. 荔枝园智能灌溉决策系统模糊控制器设计与优化[J]. 农业机械学报, 2018, 49(8): 26-32.
- [9] 赖呈纯, 黄贤贵, 王琦, 等. 果实生长与果园土壤含水量的变化对‘茂谷柑’裂果的影响[J]. 福建农林大学学报(自然科学版), 2019, 48(4): 434-439.
- [10] 邓大豪, 李得铭, 翟子翔, 等. 施用菌肥对盐胁迫下柑橘土壤理化性质和酶活的影响[J]. 海南大学学报(自然科学版), : 1-8.
- [11] 杨伟志, 孙道宗, 刘建梅, 等. 基于物联网和人工智能的柑橘灌溉专家系统[J]. 节水灌溉, 2019(9): 116-120.
- [12] 李中良, 胡晨晓, 邹腾飞, 等. 基于物联网的柑橘土壤水分养分实时监测系统的设计与实现[J]. 农业网络信息, 2014(2): 21-24.
- [13] 罗党, 王浚婷. 灰色神经网络下的多变量土壤含水量预测模型[J]. 华北水利水电大学学报(自然科学版), 2017, 38(5): 70-75.
- [14] 蔡亮红, 丁建丽. 基于变量优选和ELM算法的土壤含水量预测研究[J]. 光谱学与光谱分析, 2018, 38(7): 2209-2214.
- [15] 张武, 洪汛, 李蒙, 等. 监测采样间隔对土壤墒情预测模型性能的影响[J]. 甘肃农业大学学报, 2020, 55(1): 221-228.
- [16] 何灿隆, 沈明霞, 刘龙申, 等. 基于 NB-IoT 的温室温度智能调控系统设计与实现[J]. 华南农业大学学报, 2018, 39(2): 117-124.
- [17] 万雪芬, 崔剑, 杨义, 等. 地下 LoRa 无线传感器网络的传输测试系统研究[J]. 华南农业大学学报, 2018, 39(3): 118-124.
- [18] 岳学军, 王叶夫, 刘永鑫, 等. 基于 GPRS 与 ZigBee 的果园环境监测系统[J]. 华南农业大学学报, 2014, 35(4): 109-113.
- [19] 徐兴, 岳学军, 林涛. 基于 ZigBee 网络的水环境无线监测系统设计与实现[J]. 华南农业大学学报, 2013, 34(4): 593-597.



- [20] 余国雄, 王卫星, 谢家兴, 等. 基于物联网的荔枝园信息获取与智能灌溉专家决策系统[J]. 农业工程学报, 2016, 32(20): 144-152.
- [21] 姜晟, 王卫星, 孙道宗, 等. 能量自给的果园信息采集无线传感器网络节点设计[J]. 农业工程学报, 2012, 28(9): 153-158.
- [22] 王卫星, 陈华强, 姜晟, 等. 基于低功耗的发射功率自适应水稻田 WSN 监测系统[J]. 农业机械学报, 2018, 49(3): 150-157.
- [23] ALLEN R G, PEREIRA L S, RAES D, et al. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrig. Drain. 56[J]. Fao, Rome, 1998, 300(9): D05109.
- [24] RALLO G, GONZÁLEZ-ALTOZANO P, MANZANO-JUÁREZ J, et al. Using field measurements and FAO-56 model to assess the eco-physiological response of citrus orchards under regulated deficit irrigation[J]. Agric Water Manag, 2017, 180: 136-147.
- [25] FENG Y, PENG Y, CUI N, et al. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data[J]. Comput Electron Agric, 2017, 136: 71-78.
- [26] SPECHT D F. A general regression neural network[J]. IEEE Trans Neural Netw, 1991, 2(6): 568-576.
- [27] MAJHI B, NAIDU D, MISHRA A P, et al. Improved prediction of daily pan evaporation using Deep-LSTM model[J]. Neural. Comput. Appl., 2019, 4(6):110-122
- [28] ZHANG J, ZHU Y, ZHANG X, et al. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas[J]. J Hydrol, 2018, 561: 918-929.
- [29] CHENG H, XIE Z, WU L, et al. Data prediction model in wireless sensor networks based on bidirectional LSTM[J]. Eurasip J Wirel Comm, 2019, 2019(1): 203-212.
- [30] MAJHI B, NAIDU D, MISHRA A P, et al. Improved prediction of daily pan evaporation using Deep-LSTM model[J]. Neural Comput Appl, 2019,3(2):87-99.
- [31] REDDY D S, PRASAD P R C. Prediction of vegetation dynamics using NDVI time series data and LSTM[J]. Model Earth Syst, 2018, 4(1): 409-419.
- [32] NASH J E, SUTCLIFFE J V. River flow forecasting through conceptual models part I: A discussion of principles[J]. J Hydrol, 1970, 10(3): 282-290.
- [33] ABDULLAH S S, MALEK M A, ABDULLAH N S, et al. Extreme learning machines: A

new approach for prediction of reference evapotranspiration[J]. J Hydrol, 2015, 527: 184-195.

