

基于 HSIC 的日光温室温度传感器优化配置策略

宋成宝^{1,3}, 柳平增^{2,4*}, 刘兴华^{1,3}, 于玉婷²

(1. 山东农业大学机械与电子工程学院, 泰安 271018; 2. 山东农业大学信息科学与工程学院, 泰安 271018; 3. 山东省农业装备智能化工程试验室, 泰安 271018; 4. 农业农村部黄淮海智慧农业技术重点实验室, 泰安 271018)

摘要: 在经济成本和管理便捷性约束下, 为有效监测日光温室非均匀分布的温度信息, 需要对数量有限的传感器进行优化配置, 从而降低配置冗余, 并增强温度信息丰度。该研究以山东地区冬季日光温室为研究对象, 开展了基于希尔伯特-施密特独立性准则 (Hilbert-Schmidt Independence Criterion, HSIC) 的传感器配置策略设计及验证工作。首先, 在温室内部均匀布置了 22 个温度传感器, 采集了冬季寒冷时期 28 d 共 7 880×22 组数据, 明确了温室温度场空间分布的不均匀性及多传感器配置的基本原则和方法。然后, 利用 HSIC 准则分析了传感器之间的相对独立性, 依据信息最大化原则提出了传感器配置优先级排序算法, 并根据信息增益率及配置冗余度约束提出了传感器数量选择依据, 从而建立了日光温室传感器优化配置策略。最后, 策略验证结果表明: 相比于 22 个传感器的温度数据, 南北垂直方向上选择 S6、S9、S3 传感器时, 平均温度的 RMSE、MAPE 和信息增益率 g_r 分别为 0.24 °C、1.21%、6.70%; 东西水平方向上选择 H6、H5、H2 传感器时, 平均温度的 RMSE、MAPE 和信息增益率 g_r 分别为 0.33 °C、0.70%、9.47%。上述 2 组传感器能够有效兼顾准确获取温度整体趋势信息和增强温度信息丰度的双重目标, 试验结果证实了配置策略的有效性和可行性。所提的传感器优化配置策略可以为日光温室环境监测与调控提供理论参考和数据支持。

关键词: 日光温室; 温度; 传感器; 优化配置; HSIC

doi: 10.11975/j.issn.1002-6819.2022.08.023

中图分类号: S626.5

文献标志码: A

文章编号: 1002-6819(2022)-08-0200-09

宋成宝, 柳平增, 刘兴华, 等. 基于 HSIC 的日光温室温度传感器优化配置策略 [J]. 农业工程学报, 2022, 38(8): 200-208.

doi: 10.11975/j.issn.1002-6819.2022.08.023 <http://www.tcsae.org>

Song Chengbao, Liu Pingzeng, Liu Xinghua, et al. Optimal configuration strategy for temperature sensors in solar greenhouse based on HSIC [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(8): 200-208. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.08.023 <http://www.tcsae.org>

0 引言

日光温室是中国应用面积最广的设施类型之一, 其内部的环境温度是影响作物品质和产量的主要因素之一^[1-2]。因此, 作为微气候监测、预警、建模及控制的基本变量, 温室温度信息的全面有效监测显得尤为重要。

日光温室的主要能量来源是太阳辐射, 辐射能量沿光线入射方向呈递减趋势, 加之各边界蓄热能力及有效受热面积差异、温室空间结构不规则等因素的共同影响, 导致温室内温度呈现时空分布不均匀特点^[3], 不利于室内作物的正常生长发育, 特别是冬季寒冷时期, 温室内可能会出现局部温度过低, 造成部分作物不可逆损伤甚至死亡^[4]。上述现状说明, 利用一定数量的温度传感器监测温室内温度分布信息, 为温度预警和调控提供数据支持是十分必要的。与此同时, 日光温室是一种对生产成本较敏感的温室类型, 对传感器投入成本和生产管理便捷性要求较高。综上所述,

为了利用有限数量的温度传感器, 实现准确监测温度整体趋势信息和增强温度信息丰度的双重目标, 研究适宜日光温室的温度传感器优化配置策略是必要的。

目前, 日光温室传感器优化配置的相关研究较为缺乏, 已有的研究主要针对大型连栋温室而开展。连栋温室的温度传感器配置主要有单点配置、空间均布配置、优化配置等 3 种方式。在早期温室环境建模研究中, 国内外学者一般假定温室内温度均匀分布, 温度传感器采用单点配置方式, 以单点测量值代表整个温室的温度, 与温室内温度空间分布不均匀的实际不符^[5-7]。随着对温室内温度时空分布不均匀性的认识逐步深入, 学界开始采用多个温度传感器监测温室内的温度信息。熊迎军等提出了面向多传感器的多点数据融合算法, 对传感器数据序列进行一致性检测和三次指数平滑, 以平滑处理后的数据作为温室的整体温度^[8]。为了充分反映温室内温度信息, 宗哲英通过布设大量传感器的方式, 证实了温室内温度时空分布的不均匀性, 遗憾的是并没有进一步提出传感器优化布置策略, 传感器配置冗余度较高、经济性不佳^[9]。上述方法中多传感器采用空间均布配置方式, 不利于在温度信息丰度和传感器数量冗余两者间取得良好平衡。为了利用有限数量的传感器获取时空分布不均匀的温室内温度信息, 学界开始关注温室传感器的优化

收稿日期: 2021-12-10 修订日期: 2022-01-13

基金项目: 山东省农业重大应用技术创新项目-设施农业物联网关键设备及应用技术研究 (No.SD2019ZZ019)

作者简介: 宋成宝, 博士生, 讲师, 研究方向为农业信息化与智能化。

Email: ziwei_s1@163.com

*通信作者: 柳平增, 博士生导师, 教授, 研究方向为农业物联网与大数据、智慧农业等。Email: pzliu@sdau.edu.cn

配置研究。Lin 等^[10]利用 CFD 仿真手段研究了温度传感器的优化配置原则，提出了传感器应安装于温度及风速变化率最小的位置，而 Cheng 等^[11]采用 CFD 仿真构建了基于外部环境参数的虚拟传感器网络，代替真实的传感器对温室内部环境进行监测，大大降低了监测成本，但 CFD 仿真资源需求高、仿真用时长、温室约束条件不易遍历等原因，导致借助 CFD 仿真分析获得的传感器优化配置方案难以适应长时间动态变化的实际温室生产环境。面向连栋温室的温度监测需求，Lee 等^[12]研究了基于误差最小及信息熵最大的传感器优化布置方法，但其无法同时兼顾整体温度误差最小和温度信息丰度最大的双重目标。目前，国内外关于日光温室中温度传感器优化配置的研究较少。张军华等^[13]以典型西北日光温室作为研究对象，建立了 40 通道 PT100 温度场监测系统，通过克里金插值算法对冠层温度场进行建模，分析了冠层温度极值点的分布规律，为传感器的优化配置提供了有益参考。

传感器优化配置策略研究在桥梁损伤监测及结构稳定性等领域较为广泛，如基于目标模态矩阵的有效独立法、QR 分解法、分级免疫萤火虫算法等^[14-16]。这类研究的目的是通过有限数量传感器的优化配置更准确地监测异常信号，可为温室中传感器优化配置提供参考。国内外亦有学者开展了针对大型体育场、多区域建筑等的温度传感器优化配置研究，其采用的离群值、z-index、标准差、平均值等温度数据处理手段和灰箱-卡尔曼滤波验证方式，为日光温室温度传感器的优化布置及验证提供了有益借鉴^[17-18]。亦有学者对有限空间内空气污染物扩散监测传感器优化布置进行了研究，其采用的最大监测区域覆盖原则、弱约束高斯过程模型等，为传感器布置有效性的判断依据提供了参考^[19-21]。

针对目前缺乏有效的日光温室温度传感器优化配置策略、难以利用有限数量的温度传感器兼顾准确获取温度整体趋势信息和增强温度信息丰度的双重目标，本文构建了基于物联网技术的日光温室室内温度监测系统，采用传感器空间均布方式采集了 22 个测点的温度数据，利用希尔伯

特-施密特独立矩阵分析不同测点之间的独立性，根据测点综合独立系数进行传感器配置优先级排序，进而通过冗余性分析确定温度传感器数量，最后验证了所提出的日光温室温度传感器优化配置策略的有效性，为实际生产中日光温室室内温度的监测、预警及调控提供了技术支持。

1 材料与方法

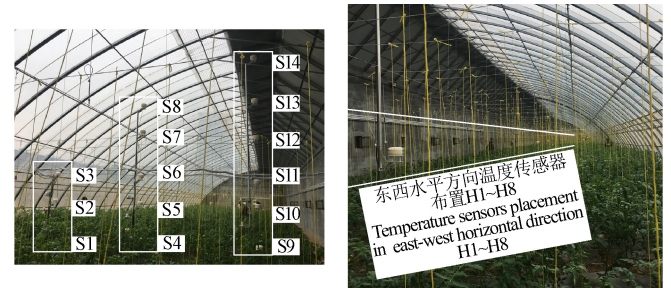
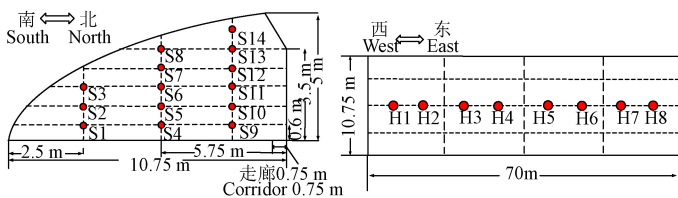
1.1 试验日光温室

试验日光温室位于山东省泰安市山东农业大学南校区蔬菜示范基地 2 号温室(东经 E117.16°, 北纬 N36.16°), 年均气温 13 °C, 温室内作物为搭架种植番茄, 温室东西长度 70 m, 南北跨度为 10 m, 脊高为 5 m, 北墙高 3.5 m, 墙体采用砖混结构, 拱架为钢架结构, 覆盖层为聚氯乙烯薄膜, 夜间覆盖保温被进行保温, 具有上下两个通风口, 冬季下通风口一般处于关闭状态, 只在上午阳光充足时打开部分上通风口进行短时通风。试验选择了典型冬季寒冷时期(2020 年 12 月 9 日至 2021 年 1 月 6 日), 采集了 28 d 的温度数据进行试验分析和策略验证。

1.2 日光温室内温度数据采集

为采集日光温室内不同区域的温度数据, 采用团队自主研发的温室环境监测系统, 包括 22 个温度传感器(DB171-10, 大连北方测控)、3 个 DTU 数据上传模块(H7710D, 深圳宏电)及物联网数据监控平台。同时, 在温室外部安装了一套监测外部气象环境的数据采集装置。监测过程中每 5 min 采集 1 次温度数据, 28 d 的监测时间内共获得 8 064×22 个样本点的数据集, 对数据进行预处理和清洗后, 每组剔除数据野 184 个, 最终获得了 7 880×22 个样本点的数据集。

课题组前期研究发现, 日光温室内南北垂直方向上的温度梯度大于东西水平方向上的温度梯度。为达到全面反映日光温室内空间温度信息的目的, 以监测点温度梯度平衡原则和等距布置原则, 在南北垂直方向上布置 14 个传感器, 东西水平方向布置 8 个传感器, 其中, 考虑到监测期内作物冠层的平均高度, 将东西方向传感器的高度设定为 1.2 m, 如图 1 所示。



a. 南北垂直方向测点布置示意图
a. Measuring points in north-south vertical direction

b. 东西水平方向测点布置示意图
b. Measuring points in east-west horizontal direction

c. 南北垂直方向测点布置现场图
c. Measuring points in north-south vertical direction in real greenhouse

d. 东西水平方向测点布置现场图
d. Measuring points in east-west horizontal direction in real greenhouse

注: S1~S14 为南北垂直方向上各测点温度传感器编号, H1~H8 为东西水平方向上各测点温度传感器编号。
Note: S1~S14 are the temperature sensor numbers of each measuring point in the north-south vertical direction, and H1~H8 are the temperature sensor numbers of each measuring point in the east-west horizontal direction.

图 1 日光温室温度传感器空间布置图
Fig.1 Spatial placement of temperature sensors in solar greenhouse

1.3 传感器优化配置策略

以传感器配置数量为约束, 为兼顾降低温室温度均值

监测误差和增强温度信息丰度的双重目标, 首先分析任意 2 个温度传感器的相对独立性, 对不同位置的传感器的优先

级进行排序, 然后根据配置冗余度约束确定传感器数量的选择依据, 最后选出满足上述目标的温度传感器组合。

1.3.1 基于希尔伯特-施密特独立性准则 (Hilbert-Schmidt Independence Criterion, HSIC) 的传感器独立性计算

温室空间中不同区域是相互联通、相互影响的, 温室内不同位置的温度传感器获取的温度数据具有较强的非线性相关性。相较于仅能反映线性相关性的皮尔逊相关系数, HSIC 是一种基于核的独立性度量方法, 具有收敛速度快、鲁棒性强、对离群值敏感性低等优点, 能够较好的反映变量间的非线性相关性^[22]。HSIC 的计算流程是在再生核希尔伯特空间 (Reproducing Kernel Hilbert Space, RKHS) 上定义 Hilbert-Schmidt 互协方差算子, 通过对该算子范数的经验估计得到独立性判断准则, 两变量的 HSIC 计算值越大, 说明两者的独立性越小, 即两者的依赖性越大; 反之亦然。因此, 本文采用 HSIC 作为温室中不同位置温度独立性的判断依据, 更加准确地反映不同监测点温度数据的相对独立性。

假设 X 、 Y 是 2 个可观测的变量, F 、 G 分别表示 X 和 Y 的再生核希尔伯特空间, 分别定义非线性映射空间 $\varphi(x) \in F$ 、 $\psi(y) \in G$, 且 $x \in X$ 、 $y \in Y$, 则相应的核函数分别记为^[23-25]

$$\begin{aligned} k(x, x') &= \langle \varphi(x), \varphi(x') \rangle, x, x' \in X \\ l(y, y') &= \langle \psi(y), \psi(y') \rangle, y, y' \in Y \end{aligned} \quad (1)$$

式中 $\langle *, * \rangle$ 表示 2 个变量的内积。

对于所有定义的正交基 $f \in F, g \in G$,

$$\langle f, C_{xy}g \rangle_F = E_{xy}([\varphi(x) - E_x(\varphi(x))][\psi(y) - E_y(\psi(y))]) \quad (2)$$

式中 $C_{xy}: G \rightarrow F$ 为交叉协方差算子

$$\begin{aligned} C_{xy} &= E_{xy}[(\varphi(x) - \mu_x) \otimes (\psi(y) - \mu_y)] \\ \mu_x &= E_x \varphi(x), \quad \mu_y = E_y \psi(y), \quad \otimes \text{为张量积。} \end{aligned}$$

将 Frobenius 范数扩展到 Hilbert-Schmidt 算子上, 得到 Hilbert-Schmidt 范数, 将该范数平方即可得到 2 个数据 x 、 y 的 Hilbert-Schmidt 独立性准则:

$$\begin{aligned} \text{HSIC}(P_{xy}, F, G) &:= \|C_{xy}\|_{\text{HS}}^2 = \sum_{f, g} \langle f, C_{xy}g \rangle_F^2 \\ &= E_{x, x', y, y'}[k(x, x')l(y, y')] + E_{x, x'}[k(x, x')]E_{y, y'}[l(y, y')] \\ &\quad - 2E_{x, y}[E_{x'}[k(x, x')]E_{y'}[l(y, y')]] \end{aligned} \quad (3)$$

式中 $\|C_{xy}\|_{\text{HS}}^2$ 为交叉协方差算子 C_{xy} Hilbert-Schmidt 范数的平方, P_{xy} 是 2 个变量 X 、 Y 的概率分布。

由上述分析可知, 在本文中 HSIC 可以有效反映任意 2 个传感器测量的温度数据之间的非线性相关性, 由此, 定义 Z 为 2 个温度变量 X 和 Y 构成的数据集

$$Z := \{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq XY \quad (4)$$

联立式 (1) (2) (3) (4), 则有:

$$\text{HSIC}(Z, F, G) = \frac{1}{n^2} \text{trace}(KHLH) \quad (5)$$

式中 $\text{trace}(\bullet)$ 为迹函数, $K, L, H \in \mathbb{R}^{n \times n}$, $K_{ij} = k(x_i, x_j)$,

$$L_{ij} = l(y_i, y_j), \quad H = I - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T, \quad \mathbf{1}_n \text{ 为 } n \text{ 维单位向量。}$$

此处, 核函数选择高斯核函数, 即

$$\begin{aligned} k(x_i, x_j) &= \exp(-\sigma^{-2} \|x_i - x_j\|^2) \\ l(y_i, y_j) &= \exp(-\sigma^{-2} \|y_i - y_j\|^2) \end{aligned}$$

N 个传感器采集的温度数据为变量 (X_1, X_2, \dots, X_N) , 利用式 (5) 可以计算不同位置温度传感器的温度数据的 HSIC 矩阵, 分析不同温度传感器的独立性, 从而为不同位置温度传感器的优先级排序提供量化依据。

1.3.2 温度传感器优先级排序算法

在利用 HSIC 量化评价不同位置温度传感器独立性的基础上, 为了给传感器优先级排序提供依据, 定义传感器个体 $s \in S$ 与传感器群体 S 中所有传感器之间综合独立系数 $r(s, S)$ 为

$$r(s, S) = \frac{1}{|S|} \sum_{s_i \in S} h(s, s_i) \quad (6)$$

式中 s 为某一个传感器个体, S 为除去 s 后传感器群体的集合, s_i 为 S 中第 i 个传感器, h 为 2 个传感器温度数据之间的 HSIC 函数, $|S|$ 为集合 S 中传感器个数。

由上述定义可知, 综合独立系数 $r(s, S)$ 越大, 传感器 s 与传感器 S 集合之间的独立性越低, 即传感器个体 s 与集合 S 中传感器群体联系越紧密, 则传感器个体 s 的温度监测值越有希望表征日光温室的总体趋势; 与之相反, $r(s, S)$ 越小, 传感器个体 s 与集合 S 中传感器群体的联系越疏远, 则传感器个体 s 越有希望提供日光温室的温度分布信息。

为兼顾准确获取温度整体趋势信息和增强温度信息丰度的双重目标, 采用如下算法确定温度传感器选择的优先级:

1) 基本参数的确定: 定义全部传感器构成的集合为 S_{tot} , 其包含的传感器个数定义为 n ; 已排序传感器集合定义为 U_{sort} , 未排序传感器集合定义为 U_{unsort} 。

2) 第 1 个传感器的选择: 选出 $r(s, S_{\text{tot}})$ 值最大的传感器 s_1 , 即传感器 s_1 与集合 S_{tot} 中传感器群体的联系最紧密, 则传感器 s_1 能够最大化反映室内温度均值, 即温度整体趋势信息。传感器 s_1 的选取如式 (7) 所示, 则集合 U_{sort} 包含第 1 个传感器 s_1 , 如式 (8) 所示。

$$s_1 = \arg \max_{s \in S_{\text{tot}}} r(s, S_{\text{tot}} \setminus \{s\}) \quad (7)$$

$$U_{\text{sort}} = \{s_1\} \quad (8)$$

式中 $S_{\text{tot}} \setminus \{s\}$ 表示 S_{tot} 中去除 $\{s\}$, U_{sort} 为已排序传感器集合, 经过一次选择后, 集合中元素个数为 1。

3) 第 i 个传感器的选择: 在集合 U_{unsort} 中选择与集合 U_{sort} 中传感器的 r 值最小的传感器为 s_i , 即传感器 s_i 与集合 U_{sort} 中已排序传感器的联系最疏远, 则传感器 s_i 可以最大化引入温度信息, 从而能更好反映日光温室内的温度分布。传感器 s_i 的选取如式 (9) 所示, 集合 U_{sort} 和 U_{unsort} 分别如式 (10) 和 (11) 所示。

$$s_i = \arg \min_{s \in U_{\text{unsort}}} r(s, U_{\text{sort}}) \quad (9)$$

$$U_{\text{sort}} = U_{\text{sort}} \cup s_i \quad (10)$$

$$U_{\text{unsort}} = S_{\text{tot}} - U_{\text{sort}} \quad (11)$$

- 4) 若 $i < n$ ，重复步骤 (3)；
- 5) 若 $i = n$ ，完成传感器优先级排序。

1.3.3 传感器数量选择依据

传感器优化配置策略应使用尽量少的传感器，实现兼顾准确获取温度整体趋势信息和增强温度信息丰度的双重目标，从而减少传感器冗余度。具体而言，在既满足所选传感器集合 U_{select} 的温度均值与全部传感器集合 S_{tot} 的均值之差低于监测误差要求，又满足所选传感器集合 U_{select} 有效表征全部传感器集合 S_{tot} 所包含的温度空间分布信息的前提下，应使传感器集合 U_{select} 包含的传感器数量最小。为实现上述目标，提出如下传感器数量选择依据。

1) 温度均值误差

根据温室温度均值监测误差要求 err （取值为 $0.5\text{ }^\circ\text{C}$ ），结合传感器优先级排序确定需要的传感器数量 n_1 ，满足其温度均值 $T_{\text{ave}-n_1}$ 与全部传感器所得温度均值 $T_{\text{ave}-\text{tot}}$ 的均方根误差（root mean square error, RMSE）小于监测误差要求，其计算如式 (12) 所示。

$$n_1 = |U_{\text{select1}}| = \arg \min_{1 \leq n_1 \leq n} (\text{RMSE}(T_{\text{ave}-n_1}, T_{\text{ave}-\text{tot}}) \leq err) \quad (12)$$

式中， U_{select1} 为满足温度均值误差阈值要求遴选出的传感器集合， $|U_{\text{select1}}|$ 为 U_{select1} 中元素的个数。

2) 温度信息丰度

传感器采集的温度数据不仅要能反映室内温度的整体趋势，而且要有足够的信息丰度，以便反映温室内的温度空间分布。信息丰度可以用温度数据的信息熵表征，熵值越大，信息丰度越大，反之亦然。Shannon 在信息论中指出离散变量 X 的熵定义为自信息的平均值，熵值计算如式 (13) 所示^[26]：

$$H(X) = -\sum_{i=1}^n p(x_i) \log p(x_i) \quad (13)$$

式中 $p(x_i)$ $i=1,2,3 \dots n$ 为变量 X 的概率分布。

2 个变量 X 、 Y 之间的条件熵定义为条件自信息的平均值，即：

$$\begin{aligned} H(Y|X) &= E [I(y|x)] = -\sum_x \sum_y p(x,y) \log p(y|x) \\ &= \sum_x p(x) [-\sum_y p(y|x) \log p(y|x)] = \sum_x p(x) H(Y|x) \end{aligned} \quad (14)$$

由式 (13)、(14) 可得，相比于原有变量 X ，新增变量 Y 所获得的信息增益为：

$$\text{IG}(Y|X) = H(Y) - H(Y|X) \quad (15)$$

相比于原温度数据集 X ，为了更加直观表示温度数据 Y 组成的新数据集 $[X, Y]$ 所获得的信息增益，采用信息增益率来度量新增传感器所获温度数据对已选传感器集合所获温度数据集的信息增益^[27-29]，信息增益率 g_r 可表示为

$$g_r = \frac{\text{IG}(Y|X)}{H(X)} = \frac{H(Y) - H(Y|X)}{H(X)} \quad (16)$$

根据式 (16) 计算温室温度试验数据发现，随着遴选

传感器集合 U_{select2} 中传感器数量 n_2 增多，新增传感器相比于集合 U_{select2} 的温度信息增益会逐渐降低。上述规律说明新增传感器对集合 U_{select2} 的温度信息丰度的贡献是越来越小的，即新增加传感器能够引入的温室温度信息是不断下降的，而传感器冗余度是不断增大的。为避免传感器冗余度过高，在保证信息丰度的前提下，计算所选的第 n_2+1 个传感器与集合 U_{select2} 的信息增益率 $g_r(s_{n_2+1}, U_{\text{select2}})$ ，若其小于阈值 $g_{r\text{threshold}}$ （取值为 10%），则认为继续增加传感器所能引入的温度信息过少，应停止继续增加温度传感器，由此确定温度传感器数量为 n_2 ，如式 (17)。

$$n_2 = |U_{\text{select2}}| = \arg \min_{1 \leq n_2 \leq n} (g_r(s_{n_2+1}, U_{\text{select2}}) < g_{r\text{threshold}}) \quad (17)$$

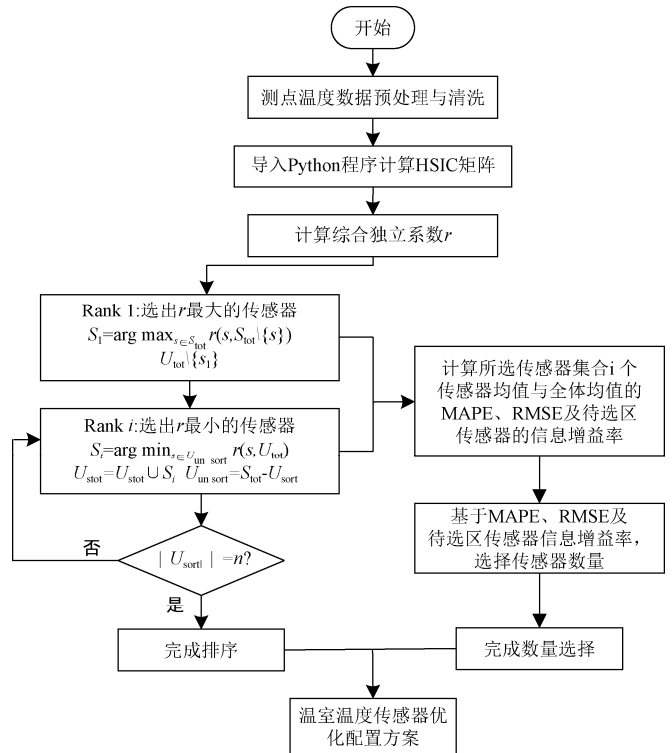
式中 U_{select2} 为满足信息增益率阈值要求遴选出的传感器集合， $|U_{\text{select2}}|$ 为 U_{select2} 中元素的个数。

3) 传感器数量的确定

为兼顾温度均值误差和温度信息丰度的双重要求，最终所需传感器数量 n_{select} 应取 n_1 和 n_2 中的较大值，如式 (18)。

$$n_{\text{select}} = \max(n_1, n_2) \quad (18)$$

传感器独立性计算、优先级排序、数量选择 3 部分内容构成温室温度传感器优化配置策略，其流程如图 2 所示。



注： S_{tot} 为全体传感器集合， n 为传感器总体数量， s 为待选传感器个体， $S_{\text{tot}} \setminus \{s\}$ 表示 S_{tot} 中去除 $\{s\}$ ， U_{sort} 为已排序传感器集合， $|U_{\text{sort}}|$ 为已排序传感器个数， U_{unsort} 为未排序传感器集合。

Note: S_{tot} is the total sensor set, n is the total number of sensors, s is the individual sensor to be selected, $S_{\text{tot}} \setminus \{s\}$ means that $\{s\}$ is removed from S_{tot} , U_{sort} is the sorted sensor set, $|U_{\text{sort}}|$ is the number of sorted sensors, U_{unsort} is the unsorted sensor set.

图 2 传感器优化配置策略流程图

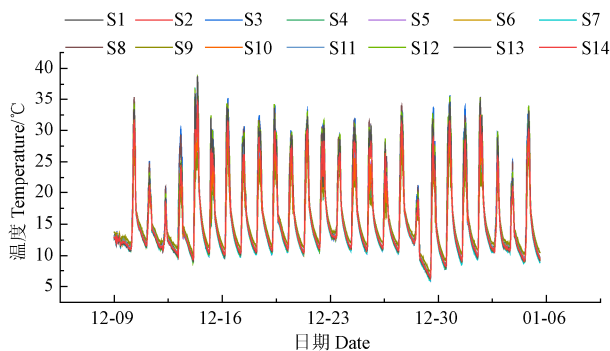
Fig.2 Flow chart of optimal sensor placement

2 结果与分析

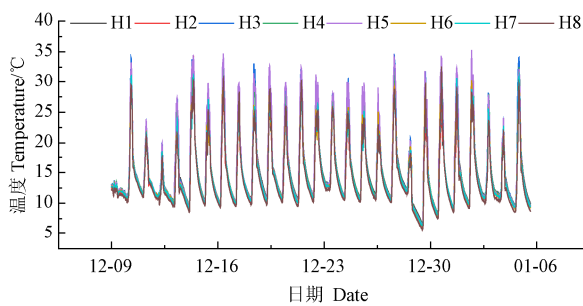
2.1 日光温室室内温度的非均匀性

2.1.1 南北垂直方向上温度分布的非均匀性

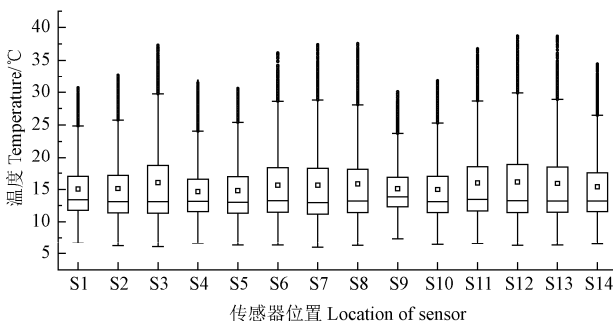
采用等距法在温室的南北垂直方向上布置了共 14 个传感器,如图 1a 所示。分析上述传感器采集的 28 d 温度数据,发现由于日光温室的能量主要来源于太阳辐射,14 个传感器的温度呈现明显的以 24 h 为单位的周期性,14 条温度曲线的趋势一致,如图 3a 所示,可初步判断不同测点的温度数据具有较强的相关性。同时,同一时刻不同测点的温度差异较大,如图 3b 所示的 1 月 4 日温度曲线,其中测点 S3 和 S9 的最大温差达 5.1 °C,温室内温度分布呈现较明显的非均匀性,并且由图 3e 可知各个测点获取的温度数据包含大量的离群值。



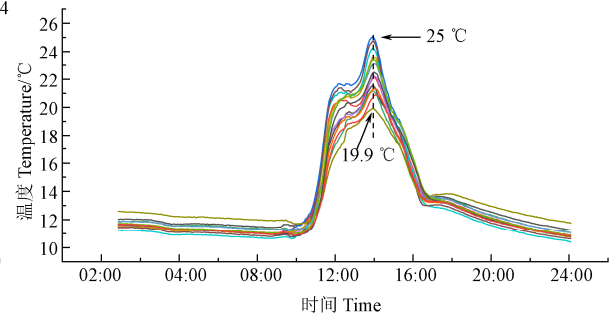
a. 垂直方向 (12-09—01-06)
a. Vertical direction (12-09—01-06)



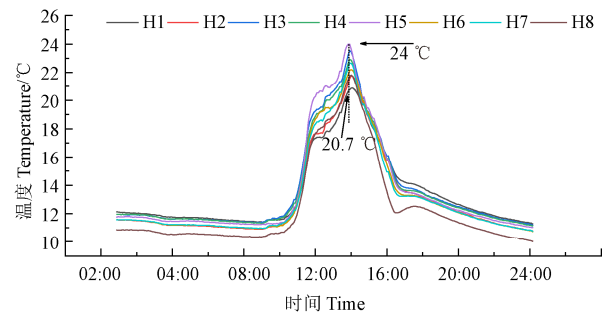
c. 水平方向 (12-09—01-06)
c. Horizontal direction (12-09—01-06)



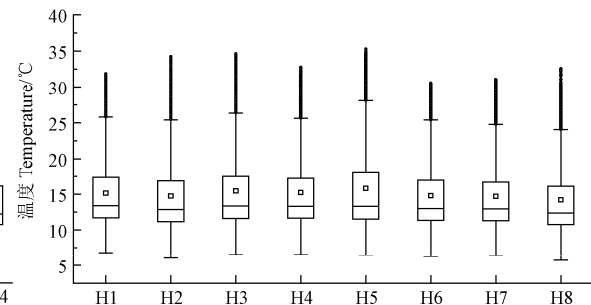
e. 垂直方向箱线图
e. Box chart in vertical direction



b. 垂直方向数据 (01-04)
b. Data in vertical direction (Jan. 4th)



d. 水平方向数据 (01-04)
d. Data in horizontal direction (Jan. 4th)



f. 水平方向箱线图
f. Box chart in horizontal direction

图 3 日光温室不同位置传感器温度

Fig.3 Air temperature of solar greenhouse measured by sensors in different positions

2.2 传感器优化配置结果与分析

基于 1.3 节的传感器优化配置策略,通过 Python 编程计算传感器之间的 HSIC 矩阵及综合独立系数 r ,利用优先级排序策略和传感器数量选择策略,可以得到以下

2.1.2 东西水平方向温度分布的非均匀性

在温室内 1.2 m 高度的东西水平方向上等距布置了 8 个温度传感器,如图 1b。分析采集的温度数据可知,水平方向的 8 个测点的温度值也呈现与南北垂直方向上类似的周期性和相关性,如图 3c 所示。同时,不同测点的温度数据也呈较大的非均匀性,如图 3d 中 H5 和 H8 的最大温差达 3.3 °C,并且 8 个测点的数据也包含大量的离群值。

综合南北垂直方向和水平方向 22 个测点的温度数据可知,温室内不同位置的温度分布是非均匀的,采用单一测点的温度数据难以充分表征室内的温度分布,采用多传感器获取不同位置的温度值是必要的。与此同时,不同测点温度的相关性说明,优化传感器的安装位置可以达到减少传感器用量、降低传感器冗余度的目标。

传感器优化配置方案。

2.2.1 传感器优先级排序

基于 1.3.2 节的优先级排序策略,可以确定南北垂直方向的 14 个传感器和水平方向的 8 个传感器的优先级排

序结果如表 1 所示。在传感器的选择过程中，依次按照优先级排序的名次进行选择，结合 2.2.2 节中均方根误差阈值和信息增益率阈值，确定传感器的数量。

表 1 南北垂直方向和东西水平方向传感器优先级排序
Table 1 Sensor ranking in north-south vertical direction and east-west horizontal direction

布置方位 Layout orientation	优先级 排序 Sensor rank	传感器 位置 Sensor location	r 值 r value	优先级排序 Sensor rank	传感器位置 Sensor location	r 值 r value
南北 垂直方向 North-south vertical direction	1	S6	646	8	S12	626
	2	S9	564	9	S1	628
	3	S3	597	10	S7	628
	4	S8	615	11	S14	631
	5	S2	619	12	S10	636
	6	S4	604	13	S11	638
	7	S13	618	14	S5	644
东西 水平方向 East-west horizontal direction	1	H6	658	5	H8	634
	2	H5	622	6	H1	640
	3	H2	622	7	H4	642
	4	H3	632	8	H7	649

注：表中 r 值为式 (6) 定义的综合独立系数。
Note: r value represents comprehensive independent coefficient in equation (6).

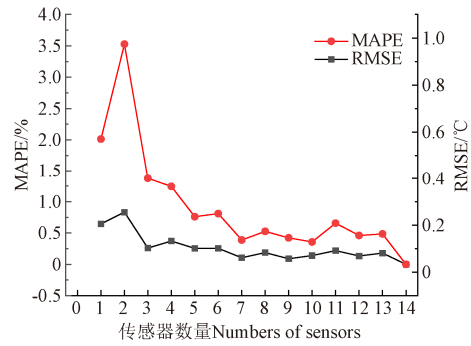
2.2.2 传感器数量的选择

随着选用传感器数量的增多，所选传感器集合与全部传感器集合的温度均方根误差 RMSE 及平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 均呈下降趋势，如图 4 所示。由图 4a 可知，南北垂直方向上传感器数量增加为 3 个时，所选传感器组合的均值与全体传感器均值之间的均方根误差 $RMSE_{vertical}$ 为 $0.26\text{ }^{\circ}\text{C}$ ，低于温度误差阈值 $0.5\text{ }^{\circ}\text{C}$ ，则依据式 (12) 得，南北垂直方向上依据误差阈值所遴选传感器的个数为 $n_{1-vertical} = 3$ 。同理，由图 4b 可知，东西水平方向上传感器数量 $n_{1-horizontal} = 3$ 时，所选传感器组合的均值与全体传感器均值之间的均方根误差 $RMSE_{horizontal}$ 为 $0.30\text{ }^{\circ}\text{C}$ ，满足阈值要求。

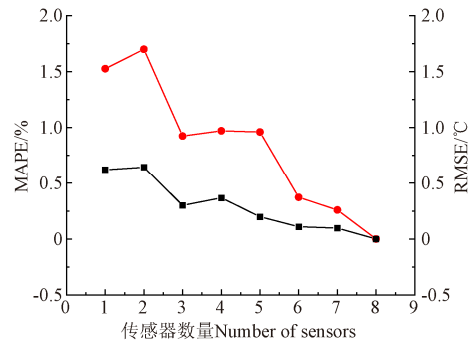
同时，新增传感器能够引入的信息增益呈现递减趋势，如图 5 所示。由图 5a 可知，南北垂直方向上传感器数量增加至 3 个时，第 3 个传感器的温度信息增益率 g_r 为 5.7%，低于阈值 10%，说明继续增加传感器能够带来的信息增益十分有限，根据式 (17) 可得，当南北方向上所选传感器为 2 个时，即可满足信息增益率要求，即依据信息增益率阈值所遴选传感器的个数 $n_{2-vertical} = 2$ 。同理，由图 5b 可知，东西水平方向上第三个传感器信息增益率为 7.3%，低于阈值 10%，因此传感器数量选择 $n_{2-horizontal} = 2$ 时，满足阈值要求。

2.2.3 传感器配置方案的确定

综合考虑温度监测误差和信息增益率要求，由式 (18) 可得，南北垂直方向和东西水平方向上所选传感器数量 $n_{vertical}$ 、 $n_{horizontal}$ 均为 3 个。结合 2.2.1 节传感器优先级排序结果，可得南北垂直方向上选择传感器 S6、S9 及 S3，东西水平方向上选择传感器 H6、H5 及 H2，综合优化配置方案如表 2 所示。



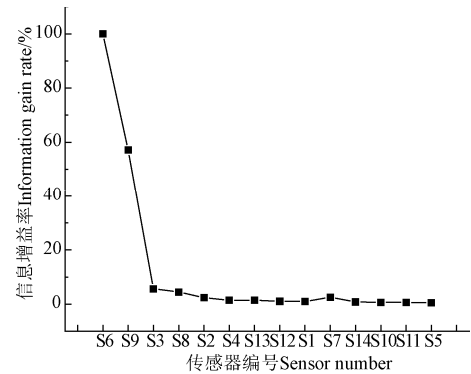
a. 南北垂直方向



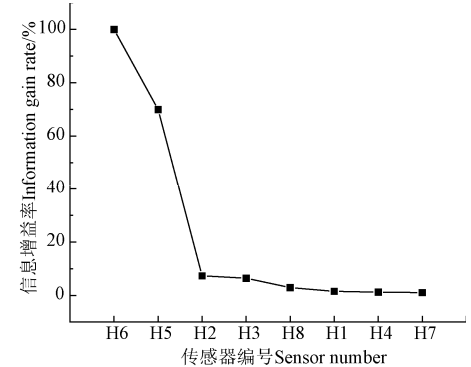
b. 东西水平方向
b. East-west horizontal direction

图 4 温度均方根误差 RMSE 及平均绝对百分比误差 MAPE 随传感器数量的变化曲线

Fig.4 The curve of Temperature Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) with number of sensors



a. 南北垂直方向
a. North-south vertical direction



b. 东西水平方向
b. East-west horizontal direction

图 5 传感器的温度信息增益率曲线

Fig.5 Information gain rate of temperature sensors

表2 温室温度传感器优配置方案

Table 2 Optimal configuration of temperature sensors in greenhouse

布置方位 Layout orientation	配置方案 Configuration
南北垂直方向 North-south vertical direction	S6、S9、S3
东西水平方向 East-west horizontal direction	H6、H5、H2
综合配置 Comprehensive configuration	S6、S9、S3、H6、H5、H2

2.2.4 传感器配置策略的验证

以2021年1月10日至17日的温度数据为例,具体分析传感器配置策略的有效性。由图6a、6b可知,当南北垂直方向、东西水平方向上分别采用(S6、S9、S3传感器组合)和(H6、H5、H2传感器组合)时,其温度均值曲线与全体传感器的温度均值曲线一致性很好, RMSE分别为0.24、0.33℃,均低于阈值(≤ 0.5 ℃), MAPE分别为1.21%和0.70%,温度信息增益率分别为6.70%和9.47%,均小于信息增益率阈值($\leq 10\%$)。上述试验数据验证了所提出的传感器优化配置策略的有效性。

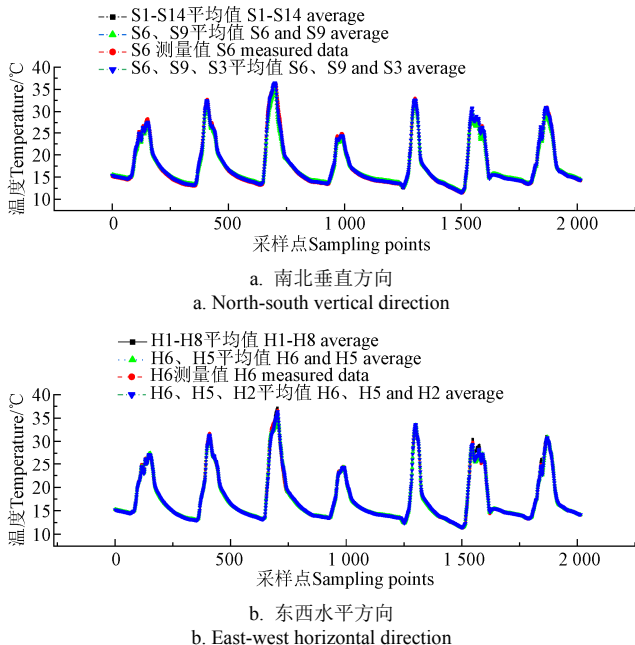


图6 不同数量传感器均值图(01-10—01-17)

Fig.6 The average temperature with number of sensors (Jan. 10th-Jan. 17th)

2.2.5 传感器配置策略讨论

分析传感器综合配置方案(南北垂直方向: S6, S9, S3; 东西水平方向: H6, H5, H2)可知:

1) 传感器空间布置特点: 根据配置方案的传感器个数选择结果和遴选顺序,结合正文图1,两个方向上均选择了作物冠层中上部的S6和H6为第1个传感器;日光温室内,作物按南北方向成行搭架种植,与这一种植模式对应,南北方向选择的第2个传感器S9位于冠层后下部,第3个传感器位于冠层前上部;东西方向选择的第2个传感器H5位于温室中间,第3个传感器位于靠近边墙部位,能更好地监测温室边缘的温度信息。

2) 作物冠层显著影响传感器的优化配置选择: 虽然本文传感器优化配置策略研究的初衷是全面准确监测温室温度场信息,但作物的光合作用、呼吸作用等过程决

定了冠层部位是温室内能量汇集、吸收、释放的主要场所;传感器配置方案中遴选的传感器主要分布在冠层的前、中、后等部位,体现了作物冠层对优化配置方案的显著影响,说明优化配置方案的构建过程中应该合理考虑作物冠层对温度场分布的影响。

3) 本文设计的优化配置策略可以较好反映作物冠层温度信息: 传感器优化配置策略在兼顾准确获取温度整体趋势信息和增强温度信息丰度的双重目标的同时,也在传感器布置时考虑了作物冠层的空间位置。优化配置结果表明所选的传感器组合可以监测冠层中部、前上部、后下部等位置的温度信息,说明所设计的优化配置策略可以为作物冠层温度监测提供足够的基础数据。

3 结论

针对北方日光温室的非均匀分布温度场的监测需要,为兼顾准确获取温度整体趋势信息和增强温度信息丰度的双重目标,本文提出了基于HSIC的温室传感器优化配置策略,通过传感器优先级排序算法和传感器数量选择算法,确定了传感器优化配置方案,最后利用试验数据验证了策略的有效性,为强成本约束下日光温室传感器的合理配置提供了决策依据。具体结论如下:

1) 由22个传感器采集的28d温度数据可知,日光温室内温度分布在和水平方向均呈现明显的非均匀性,南北垂直方向、水平方向的最大温差分别可达5.1、3.3℃。同时,传感器采集的温度数据呈现出相关性强、离群值多的特点。

2) 基于HSIC矩阵明确了不同传感器的温度数据的相对独立性,进而提出了传感器优先级排序算法和传感器数量选择依据,最终确定了南北垂直方向上选择S6、S9、S3传感器和水平方向上选择H6、H5、H2传感器的传感器配置方案。

3) 采用上述传感器配置方案时,采集的温度试验数据表明南北垂直方向的温度均值RMSE、MAPE和信息增益率 g_r 分别为0.24℃、1.21%、6.70%,水平方向的温度均值RMSE、MAPE和信息增益率 g_r 分别为0.33℃、0.70%、9.47%,均满足监测阈值要求,验证了所提传感器配置策略的有效性。

[参考文献]

- [1] 史宇亮,王秀峰,魏珉,等.日光温室土墙体温度变化及蓄热放热特点[J].农业工程学报,2016,32(22):214-221. Shi Yuliang, Wang Xiufeng, Wei Min, et al. Temperature variation, heat storage and heat release characteristics of soil wall in solar greenhouse[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(22): 214-221. (in Chinese with English abstract)
- [2] 鲍恩财,邹志荣,张勇.主动蓄热日光温室不同气流方向后墙传热CFD模拟[J].农业工程学报,2018,34(22):169-177. Bao Encai, Zou Zhirong, Zhang Yong. CFD simulation of heat transfer in back-wall of active thermal-storage solar greenhouse with different airflow directions[J]. Transactions

- of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(22): 169-177. (in Chinese with English abstract)
- [3] 蒋国振, 胡耀华, 刘玉凤, 等. 基于 CFD 的下沉式日光温室保温性能分析[J]. 农业工程学报, 2011, 27(12): 275-281.
Jiang Guozhen, Hu Yaohua, Liu Yufeng, et al. Analysis on insulation performance of sunken solar greenhouse based on CFD[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(12): 275-281. (in Chinese with English abstract)
- [4] 黎贞发, 王铁, 宫志宏, 等. 基于物联网的日光温室低温灾害监测预警技术及应用[J]. 农业工程学报, 2013, 29(4): 229-236.
Li Zhenfa, Wang Tie, Gong Zhihong, et al. Forewarning technology and application for monitoring low temperature disaster in solar greenhouses based on internet of things[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(4): 229-236. (in Chinese with English abstract)
- [5] Vanthoor B H E. A model-based Greenhouse Design Method[D]. Wageningen NL: Wageningen University, 2011.
- [6] 邹伟东, 张百海, 姚分喜, 等. 基于改进型极限学习机的日光温室温湿度预测与验证[J]. 农业工程学报, 2015, 31(24): 194-200.
Zou Weidong, Zhang Baihai, Yao Fenxi, et al. Verification and forecasting of temperature and humidity in solar greenhouse based on improved extreme learning machine algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(24): 194-200. (in Chinese with English abstract)
- [7] 刘晨霞, 马承伟, 王平智, 等. 日光温室保温被传热的理论解析及验证[J]. 农业工程学报, 2015, 31(2): 170-176.
Liu Chenxia, Ma Chengwei, Wang Pingzhi, et al. Theoretical analysis and experimental verification of heat transfer through thick covering materials of solar greenhouse[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(2): 170-176. (in Chinese with English abstract)
- [8] 熊迎军, 沈明霞, 陆明洲, 等. 温室无线传感器网络系统实时数据融合算法[J]. 农业工程学报, 2012, 28(23): 160-166.
Xiong Yingjun, Shen Mingxia, Lu Mingzhou, et al. Algorithm of real time data fusion for greenhouse WSN system[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(23): 160-166. (in Chinese with English abstract)
- [9] 宗哲英. 温室群环境参数监测系统与参数时空分布规律研究[D]. 呼和浩特: 内蒙古农业大学, 2018.
Zong Zheyang. Study on Monitoring System of Greenhouse Group and the Spatial and Temporal Distribution of the Environmental Parameters[D]. Huhhot: Inner Mongolia Agricultural University, 2018. (in Chinese with English abstract)
- [10] Lin F, Hui L, Zhi Y. Greenhouse CFD simulation for searching the sensors optimal placements[C]// 2013 Second International Conference on Agro-Geoinformatics. IEEE, 2013.
- [11] Cheng X, Li D M, Shao L M, et al. A virtual sensor simulation system of a flower greenhouse coupled with a new temperature microclimate model using three-dimensional CFD[J]. Computers and Electronics in Agriculture, 2021, 181: 105934.
- [12] Lee S Y, Lee I B, Yeo U H, et al. Optimal sensor placement for monitoring and controlling greenhouse internal environments[J]. Biosystems Engineering, 2019, 188: 190-206.
- [13] 张军华, 沈楷程, 陈丹艳, 等. 基于物联网的日光温室冠层特征温度时空变化规律分析[J]. 农业机械学报, 2021, 52(7): 335-342.
Zhang Junhua, Shen Kaicheng, Chen Danyan, et al. Spatio-temporal variation of canopy characteristic temperature in solar greenhouse[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7): 335-342. (in Chinese with English abstract)
- [14] 杨振伟, 周广东, 伊廷华, 等. 基于分级免疫萤火虫算法的桥梁振动传感器优化布置研究[J]. 工程力学, 2019, 36(3): 63-70.
Yang Zhenwei, Zhou Guangdong, Yi Tinghua, et al. Optimal vibration sensor placement for bridges using gradation-immune firefly algorithm[J]. Engineering Mechanics, 2019, 36(3): 63-70. (in Chinese with English abstract)
- [15] Brehm M, Zabel V, Bucher C. Optimal reference sensor positions using output-only vibration test data[J]. Mechanical Systems and Signal Processing 2013, 41(1): 196-225.
- [16] Kim T, Youn B D, Oh H. Development of a stochastic effective independence (SEFI) method for optimal sensor placement under uncertainty[J]. Mechanical Systems and Signal Processing, 2018, 111: 615-627.
- [17] Arnesano M, Revel G M, Seri F. A tool for the optimal sensor placement to optimize temperature monitoring in large sports spaces[J]. Automation in Construction, 2016, 68: 223-234.
- [18] Suryanarayana G, Arroyo J, Helsen L, et al. A data driven method for optimal sensor placement in multi-zone buildings[J]. Energy and Buildings, 2021, 243: 110956.
- [19] Sharma H, Vaidya U, Ganapathysubramanian B. A transfer operator methodology for optimal sensor placement accounting for uncertainty[J]. Building and Environment, 2019, 155: 334-349.
- [20] Dur T H, Arcucci R, Mottet L, et al. Weak Constraint Gaussian Processes for optimal sensor placement[J]. Journal of Computational Science, 2020, 42: 101110.
- [21] Löhner R, Camelli F. Optimal placement of sensors for contaminant detection based on detailed 3D CFD simulations[J]. Engineering Computations, 2005, 22(3): 260-273.
- [22] Wang T H, Dai X L, Liu Y Z. Learning with Hilbert-Schmidt independence criterion: A review and new perspectives[J]. Knowledge-Based Systems, 2021, 234: 107567.
- [23] Lin F, Di T, Zhang Y. HSIC-based Kernel independent component analysis for fault monitoring[J]. Chemometrics & Intelligent Laboratory Systems, 2018, 178: 47-55.
- [24] Rustamov R M, Klosowski J T. Kernel mean embedding based hypothesis tests for comparing spatial point patterns[J]. Spatial Statistics, 2020, 38: 100459.
- [25] Gretton A, Bousquet O, Smola A, et al. Measuring statistical dependence with Hilbert-Schmidt norms[C]//16th International Conference on Algorithmic Learning Theory. ALT, 2005.
- [26] 田宝玉. 信息论基础(第二版)[M]. 北京: 人民邮电出版社, 2015.
- [27] Pei X Y, Yi T H, Qu C X, et al. Conditional information entropy-based sensor placement method considering separated model error and measurement noise[J]. Journal of

Sound and Vibration, 2019, 449: 389-404.

[28] Zhang J, Lin Z W, Tsai P W, et al. Entropy-driven data aggregation method for energy-efficient wireless sensor networks[J]. Information Fusion, 2020, 56: 103-113.

[29] Wang X, Ma J J, Wang S. Parallel energy-efficient coverage optimization with maximum entropy clustering in wireless sensor networks[J]. Journal of Parallel and Distributed Computing, 2009, 69(10): 838-847.

Optimal configuration strategy for temperature sensors in solar greenhouse based on HSIC

Song Chengbao^{1,3}, Liu Pingzeng^{2,4*}, Liu Xinghua^{1,3}, Yu Yuting²

(1. College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an 271018, China; 2. College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China; 3. Shandong Agricultural Equipment Intelligent Engineering Laboratory, Tai'an 271018, China; 4. Key Laboratory of Huang-Huai-Hai Smart Agricultural Technology, Ministry of Agriculture and Rural Affairs, P.R.China, Tai'an 271018, China)

Abstract: Greenhouse has been one of the most widely passive solar energy applications for vegetable production in North China. Among them, the internal temperature is one of the most concerned environmental parameters in a greenhouse for crop growth. However, there is a non-uniform temporal and spatial distribution of temperature in the greenhouse. It is necessary to optimize the configuration with a limited number of temperature sensors for cost-saving and efficient management. In this study, an optimal configuration strategy was proposed for the temperature sensors in the greenhouse using the Hilbert Schmidt independence criterion (HSIC). The temperature was measured to reduce the redundancy of temperature sensors for the rich information. A Chinese solar greenhouse was taken as the research object, which located in Taian, Shandong Province, China. Specifically, the length, the width, the north wall, and the ridge height were 70, 10, 3.5, and 5 m, respectively. The testing period was four weeks during the coldest season from Dec. 9, 2020, to Jan. 6, 2021. Firstly, 22 temperature sensors were divided into two groups to monitor the temperature distribution in north-south vertical and east-west horizontal directions. A total of 7880×22 samples were collected in four weeks at a sampling interval of 5 min. A systematic analysis was made to clarify the non-uniformity of temperature spatial distribution in the greenhouse in the necessity of multi-sensors configuration, according to the monitoring data. Secondly, the different sensors were quantitatively evaluated for relative independence using HSIC. The sorting priority of sensor configuration was then proposed to maximize the comprehensive independent coefficients of sensors. Meanwhile, the sensor quantity was also selected to consider the constraints of sensor redundancy and information gain rate. Thirdly, the sorting and selection constituted the configuration strategy for the temperature sensors. As such, the sorting of sensors along the north-south vertical direction was ranked as the S6, S9, S3, S8, S2, S4, S13, S12, S1, S7, S14, S10, S11, and S5, according to the configuration strategy. With the increase of selected sensors according to sensors selection sequence, both the Root Mean Square Errors (RMSE) of average temperature values between the selected and all sensors and the information gain rates of the n th temperature sensor all decreased rapidly. When the number of selected sensors according to sensors selection sequence was 3, the RMSE of average temperature values between the selected and all sensors was $0.26\text{ }^{\circ}\text{C}$ and the information gain rates of the third sensor is 4.51%. At the same time, the sorting of sensors along the east-west horizontal direction was H6, H5, H2, H3, H8, H1, H4, and H7. Similarly, with the increase of selected sensors, the RMSE of average temperature values and the information gain rates also decreased rapidly. When the number of selected sensors according to sensors selection sequence was 3, the RMSE of average temperature values between the selected and all sensors was $0.30\text{ }^{\circ}\text{C}$ and the information gain rates of the third sensor is 7.31%. The results showed that the RMSE of average temperature and information gain rate were less than the thresholds ($0.5\text{ }^{\circ}\text{C}$ and 10%), respectively, fully meeting the monitoring requirements of the greenhouse temperature field, when the sensor combination was selected (S6, S9, and S3 sensors in the north-south vertical direction, and H6, H5, and H2 in the east-west horizontal direction). Finally, the temperature data for one week from Jan.10 to Jan. 17, 2021, was sampled to evaluate the effectiveness of the proposed strategy. The validation showed that the RMSE and information gain rate were $0.24\text{ }^{\circ}\text{C}$, and 6.70%, respectively, along the north-south vertical direction, and $0.33\text{ }^{\circ}\text{C}$, and 9.47%, respectively, along the east-west horizontal direction, all of which were meeting the threshold requirements. Consequently, the overall trend and information abundance of the greenhouse temperature field were optimized with the three sensors S6, S9, and S3 in the north-south vertical direction, while H6, H5, and H2 in the east-west horizontal direction, respectively, indicating better adaptability and effectiveness of the strategy. The optimal configuration strategy of temperature sensors can greatly contribute to accurately monitoring the overall trend of temperature for the higher abundance of temperature information with a limited number of sensors. The finding can provide a useful theoretical reference and technical support for the environmental monitoring and control of solar greenhouses.

Keywords: solar greenhouse; temperature; sensors; optimal configuration; HSIC