

# 珠江三角洲地区用水量影响要素及其关联规则\*

郑炎辉<sup>1,3,4</sup>, 徐小迪<sup>1,2</sup>, 李俊辉<sup>1,2</sup>, 林树彦<sup>1</sup>, 何艳虎<sup>1,2</sup>

1. 广东工业大学生态环境与资源学院, 广东 广州 510006
2. 广东省流域水环境治理与水生态修复重点实验室, 广东 广州 510006
3. 广州丰泽源水利科技有限公司, 广东 广州 510663
4. 南方科技大学环境科学与工程学院, 广东 深圳 518055

**摘要:** 区域用水量影响要素及其关联规则识别对于合理预测用水需求和优化配置水资源具有重要意义。本文基于珠三角地区历年水资源开发利用数据和经济社会发展统计数据, 利用随机森林(RF, random forest)和人工神经网络(ANN, artificial neural network)两种机器学习模型, 并综合采用SHAP(shapley additive explanations)和部分依赖图(PDP, partial dependence plots)方法, 系统识别了珠三角地区用水量影响要素及其与用水量的关联规则, 揭示了各影响要素贡献度的时空变化特征。结果表明: 用水量影响要素按重要度排序依次是GDP、人口规模、耕地面积、人均水资源量、农田实灌单位面积平均用水量、城镇人均生活用水量; ANN模型和RF模型决定系数平均值分别在0.94和0.92以上; 用水量影响要素空间上呈现中心城市以人口为主导、周边地区以耕地面积为主导的特点; 珠三角地区用水量对于人口规模和耕地面积变化的响应最为明显。研究可为珠三角地区未来用水需求预测以及水资源空间均衡配置提供科学依据与技术支撑。

**关键词:** 人工神经网络; 随机森林; 用水量; SHAP方法; PDP

**中图分类号:** TV213.4 **文献标志码:** A **文章编号:** 2097-0137(2025)02-0022-11

## On the main influencing factors of water use and their correlation rules in the Pearl River Delta region

ZHENG Yanhui<sup>1,3,4</sup>, XU Xiaodi<sup>1,2</sup>, LI Junhui<sup>1,2</sup>, LIN Shuyan<sup>1</sup>, HE Yanhu<sup>1,2</sup>

1. College of Ecological Environment and Resources, Guangdong University of Technology, Guangzhou 510006, China
2. Guangdong Key Laboratory of River Basin Water Environment Management and Water Ecological Restoration, Guangzhou 510006, China
3. Guangzhou Fengzeyuan Water Technology Co., Ltd., Guangzhou 510663, China
4. School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

**Abstract:** Identifying the factors that influence regional water use and the corresponding regulations is crucial for accurately predicting water demand and optimizing the allocation of water resources. This study collected historical data on water resource exploitation and socio-economic statistics in the Pearl River Delta (PRD) region. Two machine learning models, namely Random Forest (RF) and Artificial

\* 收稿日期: 2024-06-29 录用日期: 2024-11-27 网络首发日期: 2025-01-21

基金项目: 国家自然科学基金(52209025, 51979043); 2024年省级水资源节约与保护专项项目;  
水利部粤港澳大湾区水安全保障重点实验室开放基金(WSGBA-KJ202302)

作者简介: 郑炎辉(1990年生), 男; 研究方向: 水文水资源; E-mail: zheng\_yanhui@foxmail.com

通信作者: 何艳虎(1985年生), 男; 研究方向: 水资源优化配置; E-mail: heyanhui456@gdut.edu.cn



ZR20240217

Neural Network (ANN), were employed to systematically identify the factors affecting water use and to uncover the associated rules in the PRD region. In addition, Shapley Additive Explanations (SHAP) and Partial Dependence Plots (PDP) were applied to enhance the interpretability of the modeling outcomes. The results indicate that the factors influencing water use, in order of importance, are GDP, population size, cultivated land area, per capita water resources, water consumption for actual irrigation per unit of farmland, and urban per capita domestic water use. The average determination coefficients of ANN and RF models are above 0.94 and 0.92, respectively. Regarding water use factors, population is the dominant influence in the central cities, while cultivated land is the principal factor in the surrounding areas. Water use in the PRD region shows the most significant response to the changes in population size and cultivated land area. This research provides a scientific basis and technical support for the future prediction of water demand and the balanced allocation of water resources in the PRD region.

**Key words:** artificial neural network; random forest; water consumption; SHAP method; PDP

水是区域发展的基础资源, 随着气候变化的持续加剧与人类活动强度的不断提升, 中国的水资源短缺问题日益严峻, 这一现象在诸多研究中得到了强调(耿思敏, 2023)。不仅在干旱地区, 在许多水源丰富的地区也面临巨大的供水安全挑战(张峰等, 2018)。合理预测区域用水量在优化水资源配置、缓解水资源短缺和保障用水安全方面具有巨大的战略意义(白鹏等, 2021)。然而, 区域用水量受社会经济发展要素、自然地理和气候条件影响, 驱动和胁迫因素复杂多变(Bao et al., 2022), 且这些因素与用水量之间存在错综复杂的关联规则, 使得准确预测用水量成为一项具有挑战性的任务(朱智伟等, 2021)。

国内外学者关于区域用水量时空变化规律及其地域差异的研究已开展了大量工作, 并取得了丰富的理论成果。这些研究集中在用水量驱动机制的识别(侍孝瑞等, 2018)、用水量时空演变规律的揭示(冯小威等, 2022)以及用水量地域差异来源的分析(He et al., 2020)等方面。研究表明经济水平、城市及农业用地占比、人口规模和水资源量均会对用水量造成不同程度的影响。传统研究方法如主成分分析(张璇, 2020)、多元回归分析(陈威等, 2017)、时间序列分析(王进, 2023)和地理信息系统(姜秋香等, 2017)等在区域用水量影响要素识别、时空变化规律揭示等方面得到了广泛应用。然而, 这些方法多数为纯数学模型, 其计算繁琐、对缺失值和异常值较为敏感, 且容易过拟合和欠拟合。相较而言, 机器学习算法作为一种新兴的研究方法, 具有训练速度快、模型泛化能力强、实现简单等特点, 已经在工程应用分析(郑帅等,

2019)、医学检验(于帆等, 2023)、经济学研究(刘丽艳等, 2020)等多个领域得到了广泛应用, 在水资源评价相关研究中也证明了其良好的适用性。例如, 李宁等(2021)采用RF算法成功解析了用水量影响因素, 郑建根等(2015)利用ANN算法构建了区域水环境承载力评估模型, Villarin et al. (2019)则结合分类与回归树和RF算法建立了城市需水量的预测模型。机器学习在用水量预测方面的研究不断演进, 从传统的时间序列分析到更复杂的数据整合和模型优化, 在提高模型可解释性和应对不确定性方面取得了重要进展, 为解决实际问题提供了新的途径(Said et al., 2021)。为了提高区域用水量预测的准确性和模型模拟效率, 研究者们越来越关注机器学习所挖掘出的用水量影响要素间的关联规则(Yan et al., 2019; Wang et al., 2021)。

本文旨在基于珠江三角洲城市群历年水资源开发利用数据和经济社会发展统计数据, 综合利用RF和ANN等机器学习模型系统辨识珠三角地区用水量的影响要素, 分别采用SHAP和PDP方法揭示各影响要素对用水量贡献度的时空变化特征, 识别各影响要素与用水量变化的关联规则, 为珠三角地区水资源需求预测和管理提供技术支持和决策依据。

## 1 研究方法

### 1.1 随机森林算法

RF算法是由Breiman(2001)在2001年提出的一种基于决策树算法的集成学习算法。RF算法通过引入随机性, 能够提高模型的多样性和泛化能

力,从而在一定程度上减小过拟合的风险,同时,RF算法具有较好的可解释性和可视化性,可以用于特征选择和类别预测等领域(姚登举等,2014)。RF算法用于决策树分裂的评估指标主要包括信息增益(IG, information gain)(陈伽洛等,2019)、信息增益率(GR, gain ratio)(崔东文等,2014)、基尼指数(GI, gini index)(张虹鸥等,2006)和最小描述长度原则(MDLP, minimum description length principle)(于磊等,2016)等。相较于IG、GR和MDLP,GI能够更准确地反映特征之间的相互关系,更好地处理不平衡的数据集。因此,本文选择GI作为RF算法特征重要性分析和特征筛选的评估依据。

GI可以定义为:从数据集随机抽取两个样本而这两个样本属于不同类别的概率,这个概率反映了数据集的纯度。GI越小,说明样本数据集的纯度越高,则特征越重要,在决策树中根据该特征进行分裂的节点深度会更小,反之则说明特征不重要。其计算方法如下(陈伽洛等,2019):

$$\text{Gini}(D) = \sum_{k=1}^n p_k(1 - p_k) = 1 - \sum_{k=1}^n p_k^2,$$

$$\text{Gini}(D, A) = \frac{D_1}{D} \text{Gini}(D_1) + \frac{D_2}{D} \text{Gini}(D_2),$$

式中 $D$ 表示数据集, $p_k$ 表示从 $D$ 随机抽取的一个样本属于第 $k$ 类别的概率, $D_1$ 、 $D_2$ 分别为 $D$ 的子集,特征 $A$ 的基尼指数含义是根据 $A$ 对数据集进行划分所得到的两个集合的纯度。

## 1.2 人工神经网络

ANN是一种模拟生物神经网络运作的数学模型(俞亭超等,2004)。它由大量的神经元(或称节点)相互连接构成,每个神经元都可以接收、处理和传递信息。每个神经元之间的连接都由一个权重来衡量其强度,而这个权重会随着学习的进行而不断调整。多层感知器(MLP, multi-layer perceptron)(王之仓等,2007)、径向基函数网络(RBF, radial basis function)(王伟等,2005)和深度卷积神经网络(DCNN, deep convolutional neural network)(黄健等,2020)等是构建ANN模型的基础结构。与RBF和DCNN相比,MLP的结构较为简单,各个神经元之间全连接,权重调整规则明确,因此更容易理解和实现,且MLP可以通过随机初始化权重,并通过反向传播算法逐步调整权重,从而找到最优的权重组合,实现用水量的预测(刘洪波等,2002;Altunkaynak et al., 2018)。因此本文采用MLPRegressor构建一种基于MLP的ANN预测模型

用于珠三角区域用水量预测。

## 1.3 SHAP法

SHAP法是一种解释模型结果的方法(周扬等,2023),可以为每个样本中的每个特征变量计算出其贡献值,以达到解释模型预测结果的效果。其核心在于通过计算在合作中个体的贡献来确定该个体的重要程度。该法主要基于博弈论概念提出的Shapley值(Shapley et al., 1971),其优点是可以将参与者之间合作产生的效益分配到每个参与者身上,在机器学习领域,这意味着Shapley值是考虑多特征协同作用的结果(Niu et al., 2023)。在完成模型的训练之后,取数据集任意个特征的任意排列组合作为模型的输入进行预测,由此得到下式中的 $V(S)$ ,然后按照公式计算向量 $V$ 。SHAP法的计算公式(罗利等,2001)

$$\phi_i(V) = \sum_{S \in \mathcal{N}} \frac{(|S|-1)!(n-|S|)}{n!} [V(S) - V(S \setminus \{i\})],$$

式中 $N$ 代表参与者, $|S|$ 表示 $S$ 中的参与者数量, $n$ 为总的参与者数量, $V(S)$ 代表 $S$ 中所有参与者总的贡献, $S \setminus \{i\}$ 表示从 $S$ 中去掉第 $i$ 个参与者而得到的集合。用该式计算得到的 $V$ 是一个具有 $n$ 个分量的向量,其中的每个分量分别代表对应参与者的贡献。

## 1.4 部分依赖图

PDP是基于所选影响要素与其他影响要素不相关的假设,展现用水量与单个影响要素的依赖关系。在选用单个用水量影响要素进行分析的同时固定其他影响要素的值,通过改变该影响要素的值并记录预测模型的响应值,从而得到与响应值的关系曲线,其中每个样本生成一条响应曲线,取其平均值绘制于坐标系中即为PDP(Friedman et al., 2001)。PDP可以展示单个要素对预测结果的影响,但其不考虑要素间的交互作用,有平均效应,因而无法反映特定数据点的具体情况。

## 2 研究区概况与数据来源

珠江三角洲地区(简称珠三角),位于中国广东省中南部。该地区包括广州、深圳、珠海、佛山、江门、肇庆、东莞、惠州和中山9个城市。珠三角地理位置优越,位于珠江河口区域,地势相对平坦,其主要供水来自直接取水的河道,占总供水量的70.4%(张赵毅等,2021;何艳虎等,2023)。然而,这种供水结构面临特定挑战,包括极端干旱年份、强咸潮事件和突发水污染等,都可能对

水资源供应造成不足的影响(蔡尚途等,2008)。此外,珠江三角洲城市群是我国3大沿海城市群之一,因人口稠密、产业发达,供水保证率要求高,各城市功能定位不一和产业用水及其影响要素复杂多变,科学认识珠三角用水量影响要素及其关联规则对于合理预测该地区未来用水需求和水资源规划管理具有重要意义。

本文以珠三角9市作为研究单元,研究时段为2004—2018年。为确保数据的可靠性和完整性,采用多个官方发布的数据源,包括《广东省水资源公报》、《广东省统计年鉴》以及各地级市水资源公报。在数据预处理过程中,进行了常规的质量检查,包括异常值检测和处理,旨在消除潜在的错误,从而确保后续分析的准确性。对于数据中存在的缺失值,采用插值法(刘鹤鹏等,2023)进行填补。在本文中,使用线性插值法对2013年各市耕地面积的缺失数据进行填补。

### 3 结果与分析

#### 3.1 用水量影响要素识别

已有研究表明经济水平(侍孝瑞等,2018)、城市及农业用地占比(严婷婷等,2020)、人口规模(李宁等,2021)和水资源量(朱智伟等,2021)均会对用水量造成不同程度的影响。基于上述研究成果,结合珠三角地区经济社会发展实际,本文选取了8个用水相关的指标,分别为GDP、人口规模、人均GDP、万元GDP用水量、城镇人均生活用水量、人均水资源量、耕地面积以及农田实灌单位面积平均用水量。GDP和人口规模通常与用水量呈正相关,耕地面积关系到农业用水,而人均水资源量和农田实灌单位面积平均用水量涉及到水资源利用效率。人均GDP、城镇人均生活用水量和万元GDP用水量则可以反映城市化和工业化对用水的影响。因此,这些指标在用水量预测中被认为是重要因素。通过RF模型对指标数据进行分析,得出特征重要性顺序为:GDP、人口规模、耕地面积、人均水资源量、农田实灌单位面积平均用水量、城镇人均生活用水量、人均GDP、万元GDP用水量。表1为多次交叉验证中不同特征个数( $X$ )的模型性能参数,特征根据RF模型的结果,按照重要性顺序依次选取,当 $X=1$ 时选取GDP,当 $X=2$ 时选取GDP和人口,以此类推。

表1表明RF模型在特征消除过程中展现出稳定的性能。当 $X=3$ 时,模型平均得分表现最佳。当

$X$ 增加至6个时,模型得分的稳定性明显增强。因此,为了在保持模型性能的同时确保稳定性,应将 $X$ 控制在3~6之间。本文以最大化模型性能、增强泛化能力、避免过拟合风险为目标,将 $X$ 定为6,特征影响要素依次为GDP、人口规模、耕地面积、人均水资源量、农田实灌单位面积平均用水量、城镇人均生活用水量。这些要素的选择在一定程度上与李宁等(2021)的研究结论相似,即常住人口、第三产业生产总值和第一产业生产总值是广东省用水总量的主要影响因素。与此类似,长三角地区(冯小威等,2022)和京津冀地区(马海涛等,2017)的研究也显示出经济发展和人口增长是导致用水量变化的主要影响因素。

#### 3.2 ANN和RF在用水量模拟中的性能评估

采用交叉验证技术,将用水量和上述6个指标数据集划分为5个子集,然后反复使用这些子集进行训练和测试,采用回归决定系数作为性能评估指标,最后将所有测试结果的性能评估指标进行平均,得到最终的评估结果(图1)。

由图1可知,ANN和RF在不同子集上的性能相对稳定,平均回归决定系数分别为0.94和0.92,表明它们在不同情况下都表现出色,具有良好的稳定性和泛化能力。相比之下ANN的拟合效果更好,因此,可认为ANN模型预测用水量具有更高的可靠性,这为基于PDP识别用水量及其影响要素的关联规则提供了基础。

#### 3.3 用水量与影响要素的关联规则

基于训练得到的ANN回归模型,根据SHAP法计算样本中各影响要素的贡献,通过PDP将模型预测值对输入值的响应可视化,以挖掘用水量影响要素的关联规则。

**3.3.1 基于SHAP法的用水量影响要素贡献度时空变化分析** 图2展示了基于线性归一化特征值SHAP的结果,反映了珠三角9市各用水量影响要素与用水量之间的响应关系。不难发现,人口规模、耕地面积和用水量之间存在明显的正相关性,即随着人口规模和耕地面积的增加,用水量相应增加。城镇人均生活用水量次之,城镇人均生活用水量与用水量之间也存在正相关性,尽管它的影响相对较小。GDP与用水量之间的关联性较低,而在使用RF进行特征选择时,GDP的重要性排名靠前,这是因为RF通常只关注单个特征与用水量的关联性,而忽略了特征之间的交互作用,而使用SHAP解释ANN模型则会将特征之间的交互作

表 1 不同特征个数的模型性能参数

Table 1 Model performance parameters with different number of features

| 特征个数 | 模型性能平均得分  | 得分标准差     |
|------|-----------|-----------|
| 1    | 0.284 721 | 0.231 870 |
| 2    | 0.908 705 | 0.121 356 |
| 3    | 0.917 697 | 0.104 320 |
| 4    | 0.916 599 | 0.087 420 |
| 5    | 0.911 238 | 0.087 736 |
| 6    | 0.915 299 | 0.081 114 |
| 7    | 0.911 971 | 0.087 038 |
| 8    | 0.910 372 | 0.091 321 |

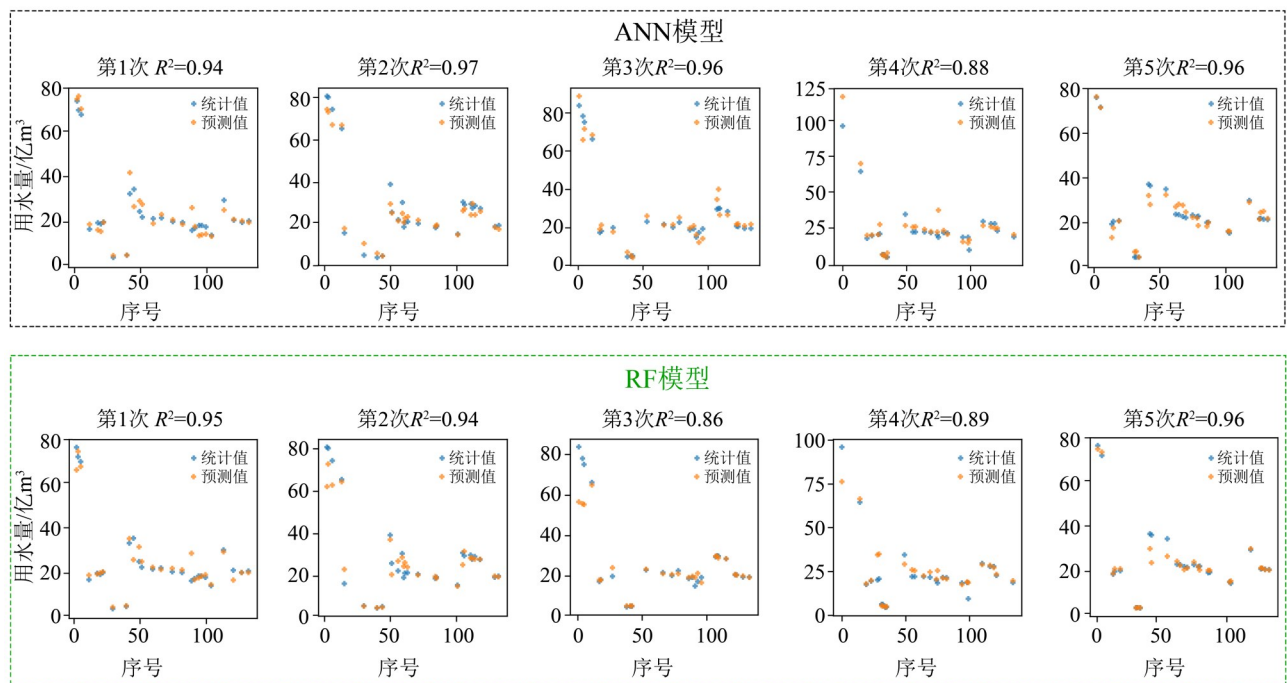


图 1 交叉验证中 ANN 模型和 RF 模型的泛化误差评估结果

Fig. 1 Evaluation results of generalization error of ANN model and RF model in cross-validation

用纳入考量。GDP 作为一个综合性的反映经济规模的指标，相对其他经济社会指标，其与用水量之间的关系较为复杂。随着 GDP 的增长，生产用水刚性需求增加，而同时用水效率也相应提高，这可从图 3 中得到佐证。用水效率的提高节省了部分生产用水需求，从而使得 GDP 与用水量具有较低的关联性。

珠三角地区不同城市各要素之间的相对贡献度各异，不同城市用水量影响要素的贡献度随时间的变化特征也各不相同(图 4)。对于广州市，在 2004—2018 年间人口规模和耕地面积对用水量的贡献度最高，其余要素对用水量的贡献度处于较低水平。对于深圳市，相对贡献度高的影响因素

有人口规模和人均水资源量，人口规模的快速增长、水资源的匮乏导致其实际用水量对人口规模、人均水资源量较为依赖。对于珠海市、佛山市、东莞市、中山市，其具有最大贡献度的影响要素分别为城镇人均生活用水量、人均水资源量、人口规模、农田实灌单位面积平均用水量。而耕地面积则为惠州市、江门市、肇庆市对用水量具有最大贡献度的影响要素。随着城市的发展，广州市耕地面积对于用水量的贡献逐年下降，直到 2018 年。相反，人口规模开始变得越来越重要，并在 2015 年超过耕地面积成为贡献度最高的影响要素。与广州市类似，深圳市的人口规模贡献度很高，并保持上升趋势，属于“人口决定型”城

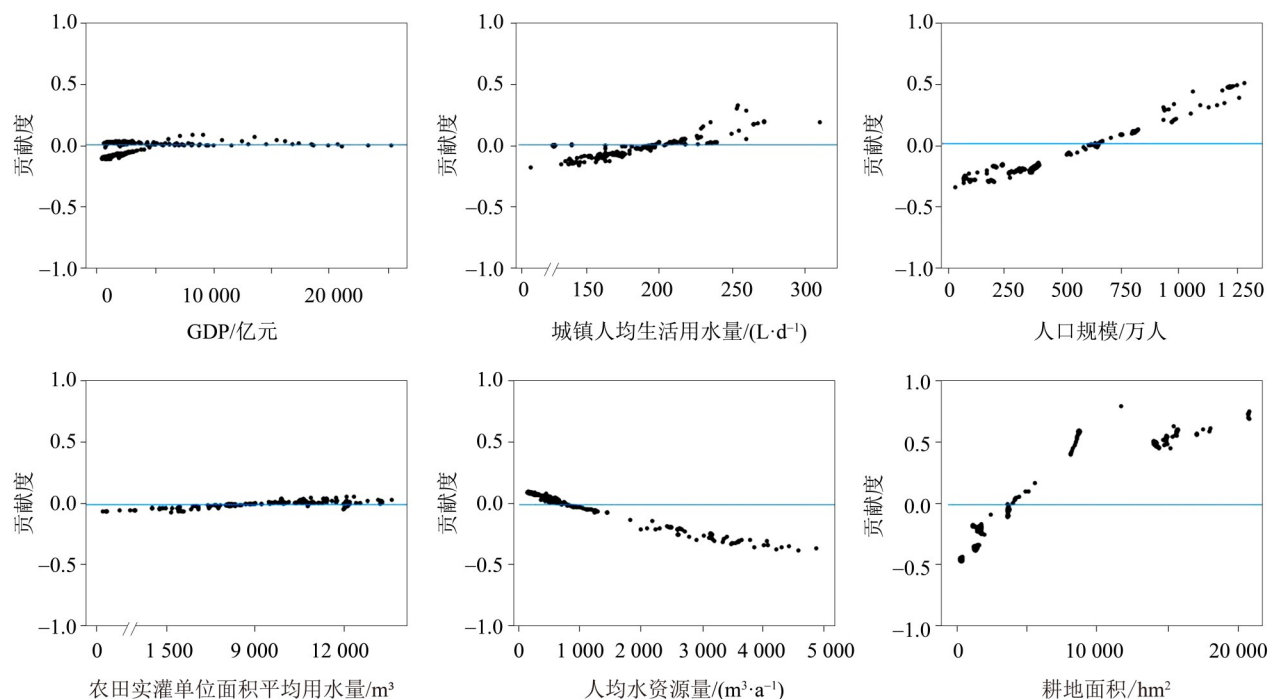


图2 影响要素的归一化特征贡献图

Fig. 2 Normalized feature contribution map of influencing factors

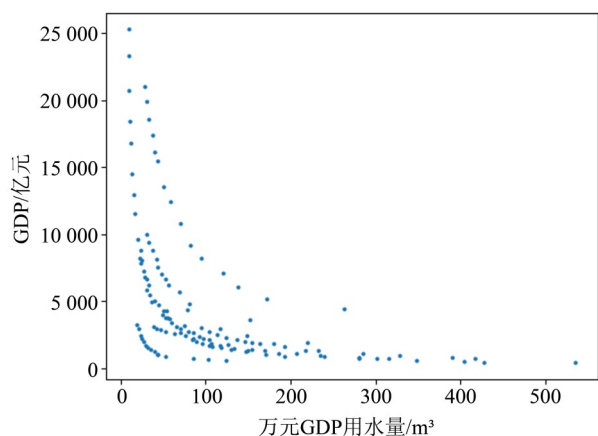


图3 GDP和万元GDP用水量关系图

Fig. 3 The relationship between GDP and water consumption per 10 000 Yuan of GDP

市。这一结论与其他特大城市的相关研究结果相似, 如上海(冯小威等, 2022)、北京(徐晋轩等, 2022)等特大城市也都显示出了人口规模主导用水量增加的趋势。惠州市、江门市和肇庆市则相反, 耕地面积的贡献度较大并保持相对稳定状态, 属于“耕地决定型”城市。在这些城市中, 农业用水占据主导地位, 这与河北(马海涛等, 2017)和成渝都市圈非中心城区(黄翱等, 2022)的农业用水占比较大的情况相类似。珠海市和东莞市的城镇人均生活用水量的贡献度较大。东莞市2010年城

镇人均生活用水量的重要性降低, 同时单位面积农田均灌溉用水也有小幅缩减。这一变化可能与人口规模的重要度上升有关。中山市2010年之前人口规模的重要度在持续上升, 同时耕地的重要性下降, 之后则趋于稳定, 这种变化可能与其产业结构升级有关。

图5呈现了各影响要素逐年贡献度在空间上的分布情况。由图5a、e和f可知, GDP、农田实灌单位面积平均用水量和城镇生活用水量的贡献度时空分布相对均匀, 且整体处于较低水平。贡献度分布均匀表明该特征具有良好的代表性, 而较低水平的贡献度意味着该特征携带的信息并非必不可少。其中, GDP反映了城市的经济规模和发展水平, 对用水量的影响很大程度上取决于产业结构、生产技术等条件。农田实灌单位面积平均用水量是一个综合性指标, 反映了农业灌溉水资源管理效能、灌溉效率与水平以及水资源利用状况等方面的内容。而城镇人均生活用水量是反映城市水资源状况和生活水平的指标, 在初期表现出相对较高的贡献度, 而后期呈现出相对均匀的时空分布。

图5b显示, 人口规模的高贡献度主要集中于珠三角地区的中心区域, 尤其是广州和深圳。这两个城市的人口规模一直具有较高的贡献度。随着与核心城市的距离逐渐增加, 人口规模的贡献

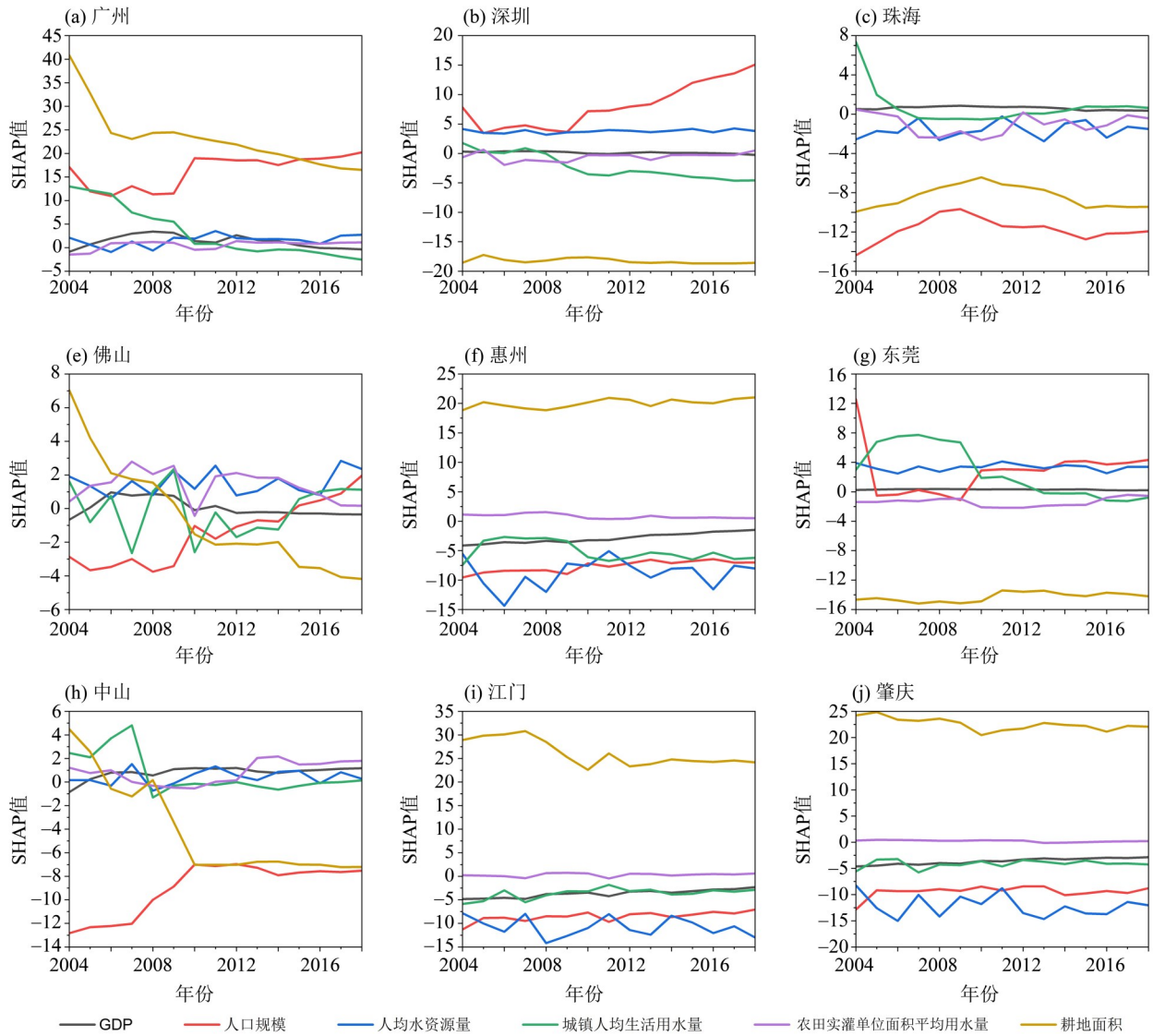


图 4 2004—2018 年珠三角 9 市用水影响要素 SHAP 值

Fig. 4 SHAP values of water use influencing factors in 9 cities in PRD from 2004 to 2018

度逐渐减小, 呈现出向外围递减的趋势。这种时空分布格局相对稳定, 突显了核心城市的人口规模对用水量的持续影响, 反映了核心城市在吸引人口方面的优势, 表明在制定水资源管理政策时需要考虑到核心城市的特殊地位。

从图 5c 可知, 耕地面积的高贡献度主要集中在珠三角地区周边(惠州市、江门市和肇庆市)。这些地区的耕地面积贡献度呈现逐年下降的趋势。这一分布特点与 Du et al.(2013)的研究结果一致。在 1990—2010 年间, 珠三角地区经历了快速的城市扩张和城镇化过程, 导致耕地面积逐年下降, 这对于水资源管理和决策提出了新的挑战, 需要综合考虑城市化发展和农业用水需求。

图 5d 表明, 人均水资源量的贡献度随时间变

化不大, 但在空间上存在显著的差异。在中心区域, 如广州、深圳、东莞和佛山, 人均水资源量对用水量的贡献度明显高于外围区域, 表明中心区域的实际用水量取决于供水侧, 随着供水量的增加此区域的用水量也会相应增加, 即中心区域实际用水量对水资源供给量的依赖程度较高。这也受到中心区域产业密集、人口聚集等因素的影响。

综合来看, 珠三角地区周边区域的耕地面积对用水量的贡献度明显高于中心区域, 说明周边区域的用水量相较于中心区域对耕地面积的依赖程度较高, 中心区域的人口规模和人均水资源量的贡献度明显高于周边地区, 说明珠三角地区用水影响要素呈现出中心城市以人口规模为主导、

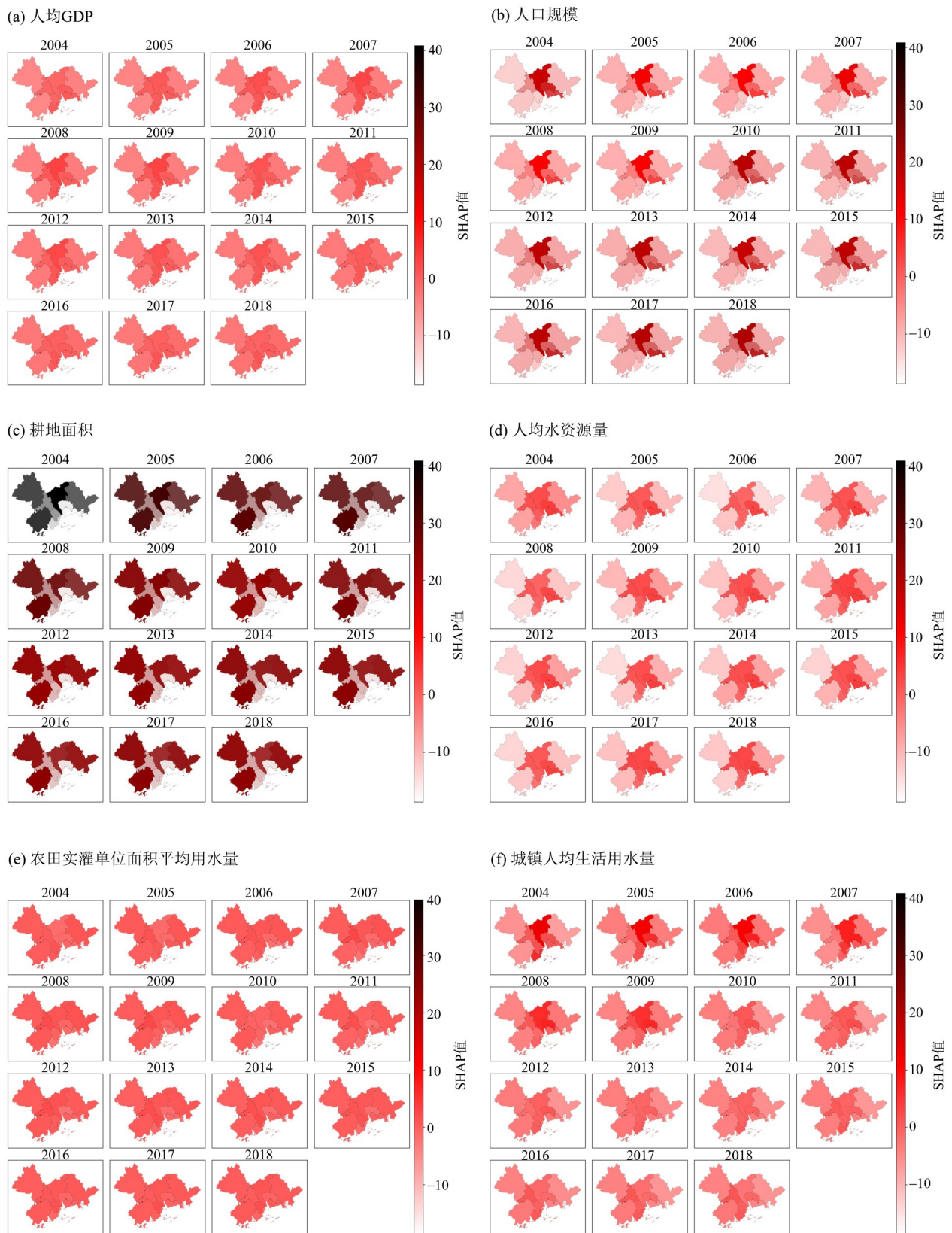


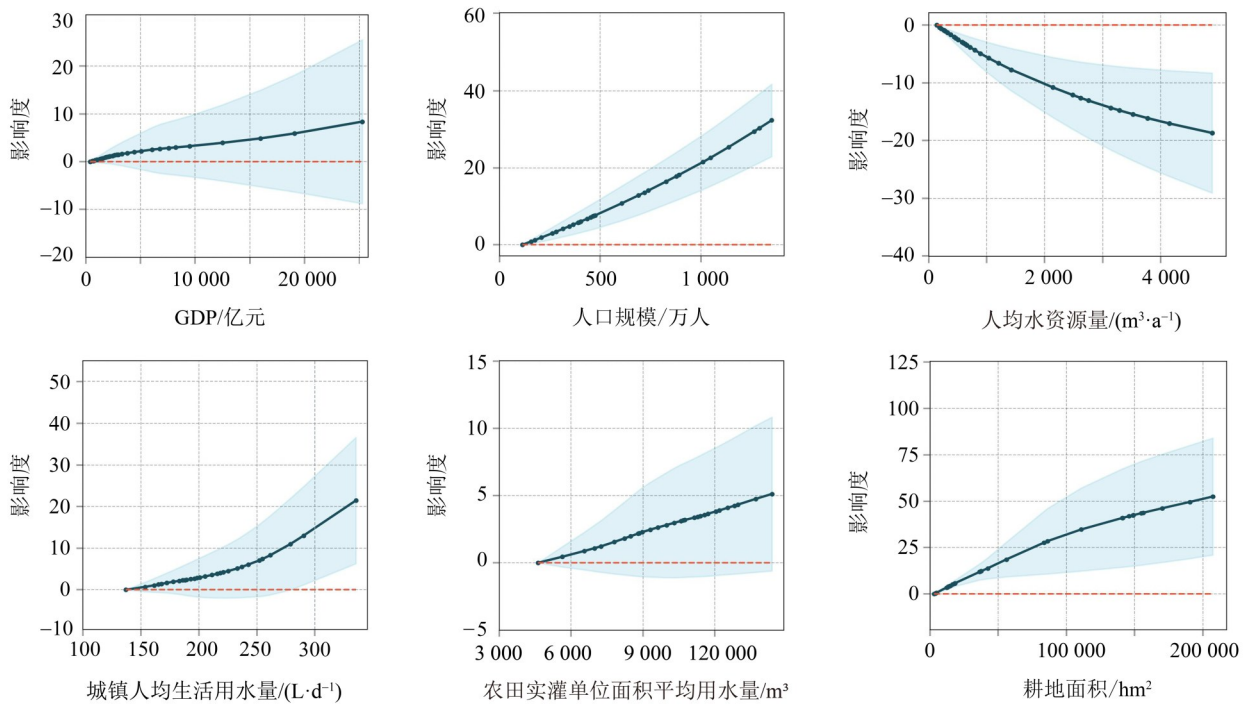
图 5 2004—2018年用水量影响要素 SHAP 值空间分布

Fig. 5 Spatial distribution of SHAP values of water use influencing factors during 2004–2018

周边地区以耕地面积为主导的特点。这与前文分析广州市、深圳市用水影响要素为“人口决定型”城市，惠州市、江门市和肇庆市为“耕地决定型”城市一致。与此同时，在经济水平相当的长三角城市群，城市用水空间分布的动态变化呈现出类似的特点，核心城市的用水量随人口增长持续保持较高水平，用水重心整体向东南人口集聚方向偏移(冯小威等,2022)。随着珠三角和长三角等地区的城市化进程不断推进，核心城市人口规模以及周边城市耕地面积等影响要素的变化值得关注。通过深入分析这些要素的变化，能够更科学地制定区域水资源管理策略，以适应区域发展的实际需求，确保水资源的可持续利用和生态安全。

**3.3.2 基于部分依赖图的用水量影响要素关联规则识别** 珠三角地区用水量影响要素的PDP (图6)，显示了在95%置信度下用水量预测结果在不同影响要素上的平均变化状况。除人均水资源量以外，随着各影响要素值的增长，用水量均呈现

出增加的趋势。从变化幅度上看，耕地面积和人口规模对结果的影响明显高出其他特征。当耕地面积从2 844 hm<sup>2</sup>增加至207 123 hm<sup>2</sup>时，对应用水量变化值高达51亿 m<sup>3</sup>，而人口从113万增加至1 350万时，对应用水量变化值也达到32亿 m<sup>3</sup>，这两个特征的变化会对用水量造成巨大影响，与前文研究结果一致。GDP对用水量的影响曲线有轻微变陡趋势，之后基本上是一条直线，这反映出经济发展和生活水平提高往往伴随着对水资源需求的增加，然而随着技术的进步，这种增长趋势可能会放缓或呈现平稳状态。而人均水资源量的负贡献可能是由模型预测用水量对人均水资源量变化的响应造成的。相较于耕地面积和人口规模，GDP、农田实灌单位面积平均用水量和城镇人均生活用水量对用水量造成的影响较弱，而越弱的影响要素越容易受与之相关的要素影响，从而被错误地解释，因此将人口和耕地面积作为用水量变化的解释是比较具有说服力的。



蓝色区域代表95%置信区间预测值,深蓝色的线为平均值,红色虚线为0点基准线

图6 用水量影响要素部分依赖图

Fig. 6 PDP of factors influencing water consumption

## 4 结 论

本文基于珠三角地区历史用水量和经济社会

统计数据，综合利用RF和ANN模型，分别采用SHAP法和PDP，识别了珠三角地区用水量与影响要素的关联规则，并分析了各影响要素对用水量

贡献度的时空变化特征。主要结论如下:

1) 用水量影响要素的重要性排序为GDP、人口规模、耕地面积、人均水资源量、农田实灌单位面积平均用水量、城镇人均生活用水量。ANN模型和RF模型模拟珠三角地区用水量的决定系数平均值分别在0.94和0.92以上,ANN模型具有较高的模拟精度。

2) 珠三角地区用水量影响要素在空间上呈现出中心城市以人口规模为主导、周边地区以耕地面积为主导的特点,惠州、江门和肇庆的用水量主要受耕地面积的影响,而广州、深圳和东莞的用水量则主要受人口规模的影响;在时间上整体呈现出耕地面积的影响逐渐减弱,而人口规模的

影响逐渐增强,尤以广州和深圳最为明显。

3) 除人均水资源量以外,各影响要素对珠三角地区用水量均具有正向促进作用,其中人口规模和耕地面积对用水量变化的影响最大,其次为城镇人均生活用水量、GDP和农田实灌单位面积平均用水量。

随着用水大数据监测技术和计算机算力的提升,采用机器学习模拟用水量有望在提高用水量预测准确性和稳定性方面取得更大的突破。为了更好地满足水资源管理和决策的需求,未来研究可将基于数据驱动的机器学习模型与基于机理机制的需水量预测模型相结合,进一步增强用水量预测的合理性与可靠性。

## 参考文献:

- 白鹏,龙秋波,2021.3种用水量预测方法在京津冀地区的适用性比较[J].水资源保护,37(2):102-107.
- 蔡尚途,吴怡蓉,张虹,2008.珠江压咸补淡水调度以人为本确保珠三角供水安全[J].中国水利,(24):44-47.
- 陈伽洛,陈龙然,2019.决策树与随机森林[J].信息与电脑,31(17):43-45.
- 陈威,艾婵,2017.基于多元线性回归模型的武汉市水资源承载力研究[J].河南理工大学学报(自然科学版),36(1):75-79.
- 崔东文,金波,2014.基于随机森林回归算法的水生态文明综合评价[J].水利水电科技进展,34(5):56-60+79.
- 冯小威,樊良新,王宝山,等,2022.基于夜间灯光数据的长三角城市群城市用水时空演变及驱动研究[J].水资源与水工程学报,33(4):97-107.
- 广东省水利厅,2005—2019.广东省水资源公报[R].北京:中国水利水电出版社.
- 广东省统计局,2005—2019.广东统计年鉴[M].北京:中国统计出版社.
- 耿思敏,2023.从国内外用水效率看中国节水空间[J].水利发展研究,23(11):78-83.
- 何艳虎,龚镇杰,吴海彬,等,2023.粤港澳大湾区城市生态效率时空演变及影响因素[J].生态环境学报,32(3):469-480.
- 黄翱,甘治国,常文娟,等,2023.成渝地区双城经济圈用水效率及节水潜力研究[J].水利水电技术(中英文),54(9):74-85.
- 黄健,张钢,2020.深度卷积神经网络的目标检测算法综述[J].计算机工程与应用,56(17):12-23.
- 姜秋香,周智美,王子龙,等,2017.基于水土资源耦合的水资源短缺风险评价及优化[J].农业工程学报,33(12):136-143.
- 李宁,汪丽娜,2021.基于随机森林回归算法的用水总量影响因素解析——以广东省为例[J].华南师范大学学报(自然科学版),53(1):78-84.
- 刘鹤鹏,李登华,丁勇,2023.融合多测点数据相关性的大坝监测历史数据填补[J].人民长江,54(9):245-251.
- 刘洪波,张宏伟,田林,2002.神经网络法预测时用水量[J].中国给水排水,18(12):39-41.
- 刘丽艳,朱成全,2020.机器学习在经济学中的应用研究[J].天津师范大学学报(社会科学版),(2):51-58.
- 罗利,鲁若愚,2001. Shapley 值在产学研合作利益分配博弈分析中的应用[J].软科学,15(2):17-19+73.
- 马海涛,耿凤娟,2017.京津冀区域城镇化过程中水资源利用的影响因素[J].科学,69(5):32-36.
- 侍孝瑞,王远坤,卞锦宇,等,2018.水资源承载力关键驱动因素识别研究[J].南京大学学报(自然科学版),54(3):628-636.
- 王进,2023.基于时间序列分析的阿克苏地区用水预测研究[J].地下水,45(4):136-137.
- 王炜,吴耿锋,张博锋,等,2005.径向基函数(RBF)神经网络及其应用[J].地震,25(2):19-25.
- 王之仓,邓伟,2007.多层感知器学习算法研究[J].青海师范大学学报(自然科学版),(1):37-39.
- 徐晋轩,杨默远,潘兴瑶,等,2022.基于水足迹的北京市水资源开发利用演变特征分析[J].水资源保护,38(5):96-104.
- 姚登举,杨静,詹晓娟,2014.基于随机森林的特征选择算法[J].吉林大学学报(工学版),44(1):137-141.
- 于帆,何海洪,周义文,2023.人工智能在检验医学领域的应用进展[J].国际检验医学杂志,44(18):2267-2273.

- 于磊,王普,赵寒,等,2016. 基于 MDLP-Apriori 算法的离散 Shannon 熵值标签排序[J]. 计算机应用研究, 33(6): 1633-1636.
- 俞亭超,张土乔,毛根海,等,2004. 预测城市用水量的人工神经网络模型研究[J]. 浙江大学学报(工学版), 38(9): 1156-1161.
- 严婷婷,刘定湘,颜霜,等,2020. 黄河流域“以水而定,量水而行”的思考与建议——基于用水变化及驱动因素分析[J]. 水利发展研究, 20(10): 54-58.
- 张峰,薛惠锋,罗婷,2018. 广东省水资源短缺指数构建与驱动测度——基于物元关联模型[J]. 北京理工大学学报(社会科学版), 20(1): 52-61.
- 张虹鸥,叶玉瑶,陈绍愿,2006. 珠江三角洲城市群城市规模分布变化及其空间特征[J]. 经济地理, 26(5): 806-809.
- 张璇,2020. 基于 PCA-BP 神经网络的郑州市年用水量预测[J]. 科学技术创新, (28): 107-111.
- 张赵毅,何艳虎,林柱良,等,2021. 基于 Copula 函数的珠江三角洲上游来水丰枯遭遇分析[J]. 人民珠江, 42(12): 30-41.
- 郑建根,胡荣祥,2015. RBF 神经网络在区域水环境承载力研究中的应用[J]. 中国农村水利水电, (9): 52-54.
- 郑帅,姜谔男,张峰瑞,等,2019. 基于机器学习与可靠度算法的围岩动态分级方法及其工程应用[J]. 岩土力学, 40(S1): 308-318.
- 周扬,邵天元,钱才云,2023. 南京市城市轨道交通站点周边地区建成环境对居民活动的影响——基于梯度提升决策树与 SHAP 解释模型的分析[J]. 科学技术与工程, 23(17): 7509-7519.
- 朱智伟,闫文晶,张海涛,等,2021. 基于大数据分析的郑州市城市用水量预测[J]. 华北水利水电大学学报(自然科学版), 42(4): 26-30.
- ALTUNKAYNAK A, NIGUSSIE T A, 2018. Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models[J]. Urban Water J, 15(2): 177-181.
- BAO Q L, DING J L, HAN L J, et al, 2022. Predicting land change trends and water consumption in typical arid regions using multi-models and multiple perspectives [J]. Ecol Indic, 141: 109110.
- BREIMAN L, 2001. Random forest[J]. Mach Learn, 45: 5-32.
- DU S, SHI P, ROMPAEY V A, 2013. The relationship between urban sprawl and farmland displacement in the Pearl River Delta, China[J]. Land, 3(1): 34-51.
- FRIEDMAN J H, 2001. Greedy function approximation: a gradient boosting machine[J]. Ann Stat, 29(5): 1189-1232.
- HE Y, LIN Z, CHEN X, 2020. Regional difference of water use in a significantly unbalanced developing region [J]. Water Policy, 22(6): 1182-1199.
- NIU X, LU C, ZHANG Y, et al, 2023. Hysteresis response of groundwater depth on the influencing factors using an explainable learning model framework with Shapley values [J]. Sci Total Environ, 904: 166662.
- SAID N M, ZIN Z M, ISMAIL M N, et al, 2021. Univariate water consumption time series prediction using deep learning in neural network (DLNN) [J]. Int J Adv Technol Eng Explor, 8(76): 473-483.
- SHAPLEY L S, SHUBIK M, 1971. The assignment game I: The core[J]. Int J Game Theory, 1(1): 111-130.
- VILLARIN M C, RODRIGUEZ-GALIANO V F, 2019. Machine learning for modeling water demand[J]. J Water Resour Plan Manag, 145(5): 04019017.
- WANG F, WANG Y, ZHANG K, et al, 2021. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation [J]. Environ Res, 202: 111660.
- YAN J, JIA S, LV A, et al, 2019. Water resources assessment of China's transboundary river basins using a machine learning approach[J]. Water Resour Res, 55(1): 632-655.

(责任编辑 秦社彩)