利用环境与电气测量数据，通过人工神经网络估计光伏板的功率产出

Antony Morales-Cervantes、Oscar Lobato-Nostroza、Gerardo Marx Ch´avez-Campos、Yvo Marcelo Chiaradia-Masselli 和 Rafael Lara-Hern´andez

arXiv:2305.01848v1 [stat.AP] 3 May 2023

摘要——**天气是实施光伏板系统预测时面临的主要问题之一，因为天气是电能扰动和中断的主要诱因。因此，有必要选择可靠的预测模型，以更好地利用能源。本研究构建了一个测量原型，用于现场采集电压和电流测量数据，以及辐射、温度和湿度等环境因素数据。随后，对变量进行了相关性分析，并采用人工神经网络进行系统预测。采用 3 个变量（光照、温度和湿度）得到的估计效果最佳，误差为 0.255。这些结果表明，可以对光伏板系统进行良好的估计。**

索引词—**光伏发电系统；储能系统；辐射预测；人工神经网络。**

I. 引言

如今，能源消耗每年都在显著增长。因此，能源公司在生产补充能源时使用化石燃料，导致污染加剧[1]。为减轻影响并满足能源需求，企业已将替代能源纳入其能源结构 [2]。其中最重要的能源之一是太阳能，如今已成为全球最受欢迎的替代能源 [3]。多年来，人们一直利用光伏（PV）面板将太阳能转化为电能 [4],。光伏电池产生的电流和功率取决于外部因素（如环境条件）以及光伏系统的典型内部因素。

具体而言，天气会对光伏电池的电功率造成扰动和中断 [5], [6],[7], [8]。由于这种变化性，在部署这些光伏系统时，有必要实施可靠的预测，以避免因计划能源与实际产能之间的差异而受到处罚 [9]。可根据待分析的参数和具体需求选择不同的预测模型 [10]。

近年来，人工智能（AI）的应用显著增加，这得益于其对复杂计算任务进行建模和求解的能力[11]。事实上，数据采集系统中的人工智能算法有助于提高测量设备的盈利能力 [12] [13]。预测模型也是检测对光伏系统及其配电系统执行维护的适当时机的重要指标。光伏能源预测有不同的方法。超短期方法（从几秒到几分钟）用于微电网中电力市场内光伏系统的控制与管理。短期方法（48–72小时）用于能源系统运行控制、经济调度和机组组合等。中期方法（几天到一周）和长期方法（几个月到一年或更长时间）用于规划光伏系统 [14]。

因此，人们已经采用多种模型和方法来估算光伏系统的发电量。先进方法涵盖多种人工智能和机器学习技术，例如人工神经网络（人工神经网络）、k近邻（kNN）、极限学习机（ELM）和支持向量机（SVM）；这些是最常用的方法 [15]。机器学习算法可以

本文源自研究项目 6127.17-P，该项目由“墨西哥国家技术研究院”资助，并包括由赞助方开展的电子制造项目 National Laboratory SEDEAM。Gerardo Marx Ch´avez-Campos 任职于墨西哥国家技术研究院莫雷利亚技术研究所，地址为 Av. Tecnol´ogico 1500, Col. Santiaguito, CP 58120, 墨西哥；他是通讯作者：gmarx cc@itmorelia.edu.mx

分为三大类：（1）1）监督学习，算法在这种学习方式中建立输入和输出特征之间的关系。2）无监督学习，算法在这种学习方式中寻找模式和规则，以更好地描述数据。3）强化学习，主要用于处理大量数据，并对其进行精简，以便于可视化或分析 [16]。

光伏系统的功率预测主要基于人工神经网络，因为涉及的参数十分复杂。人工神经网络旨在模拟人脑的行为，并生成用于决策的响应 [17]。每个人工神经网络的基本组成部分是其处理单元——神经元。神经网络通过不同的方法组织这些处理单元，以满足不同的数值处理需求 [18]。

近期报道的预测模型均以人工神经网络为基础。一类模型使用人工神经网络，根据过往气象数据（温度、湿度和降雨概率）及辐射，估算（特定城市的）太阳辐射；再利用估算出的辐射，尝试预测光伏系统的输出能量 [19], [20], [21]。其他模型则使用人工神经网络，直接根据不同输入估算发电量，例如：（i）湿度 [22]；（ii）温度、湿度和降雨量 [23]；（iii）太阳辐射和温度 [24]；（iv）太阳能功率和气象数据 [25]。此外，一种改进模型会结合数据相关性和人工神经网络，筛选出最关键的输入 [26]。然而，大多数模型都基于可在线获取的数据库，或并非在光伏系统所在位置精确测得的气象数据。因此，某个区域的数据资料并不充分，有时也无法获得连续测量数据。另一方面，众所周知，半导体对高温极其敏感。

因此，本文提出了一种物联网设备，用于测量和记录光伏系统的现场数据。在为期—120 天—的测量期间，以 5 分钟的采样频率采集了太阳辐射、太阳能电池板的温度和湿度，以及电池板的电功率（电压和电流），共获得超过 32,200 条测量数据。随后，利用这些测量数据训练了多种人工神经网络拓扑结构，并将其与多元线性回归器模型进行了比较。最佳拓扑结构的误差水平为 0.255326464，表明该数据模型具有可靠性。

II. 材料与方法

本研究方法包括采集并验证光伏系统的数据，以便进行数据分析，并利用人工神经网络预测电功率（见图 1）。因此，我们设计了一种测量原型，用于采集现场电压和电流测量值，以及辐射水平、温度和湿度等环境因素。最终，通过分析原型的读数、执行验证流程并分析变量间的相关性，获得系统的运行特性。随后，利用训练集建立回归模型，最终实现人工神经网络。我们使用均方根误差（RMSE）作为主要指标，通过公式 1 将所得估计值与测试集进行比较评估。

$$RMSE = \sqrt{\frac{\sum (x_T - x_i)}{N}} \tag{1}$$

其中，x_T 为估计值，x_i 为实际值，N 为测量总数。

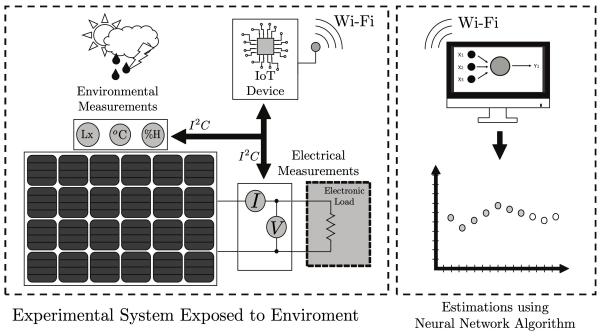


图 1：用于数据采集与分析的实验装置。

A. 实验装置

数据通过一套嵌入了三个带有 I^2C 通信的传感器的独立物联网系统进行采集。该物联网系统还嵌入了一个Web服务器，用于配置和管理记录数据[27]。图2展示了由三个主要部分组成的实验系统。最左侧

包括环境变量传感器、OPT3001和HDC2080集成电路。OPT3001是一款光传感器，分辨率为0.01 lx，上限为128 klx。然而，为将设备的量程扩展55%，使用了衰减玻璃；随后使用商用数字照度计MASTECH ms6612进行了校准程序。HDC2080传感器测量相对湿度（RH）和温度。对于温度，该传感器的分辨率为 ±2°C，测量范围为 −40°C至85°C；对于RH，该传感器提供分辨率为 ±0.2%的测量结果。

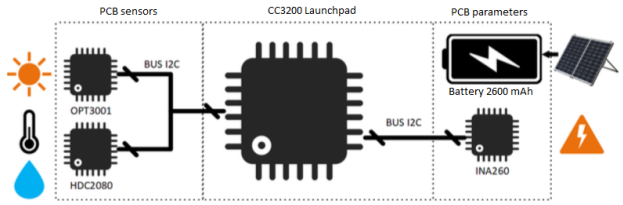


图2：用于测量的实验装置系统。

B. 日志数据分析

在进行预测过程之前，已将记录数据与类似的商用测量系统进行比较，以评估性能。根据中心极限定理，我们的数据呈正态分布，因此采用皮尔逊相关法对实测变量之间的线性依赖关系进行测量，以确定其在预测过程中的重要性。

将该行为与莫雷利亚技术学院气象站的数据进行比较。两组数据均由太阳辐射、湿度和温度组成，对应于2018年11月12日至2018年12月6日期间。由于测量数据来自不同设备，且设备的量程和位置不同（相距5米），因此使用公式2对测量数据进行了归一化：

$$x = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{2}$$

其中，x_i 是待归一化的实际值，x_{min} 是整个数据集的最小值，x_{max} 是最大值。图 3 显示了物联网系统和气象站仅针对辐照度变量采集的测量数据。数据集之间的细微差异可归因于各系统所在的位置；不过，整体性能与此相关。

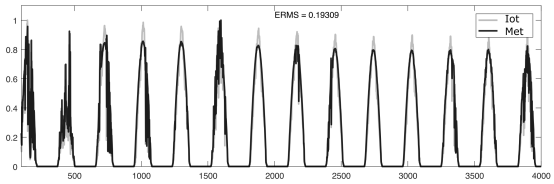


图3：气象站与物联网系统之间的数据比较。

图4展示了三幅相关性图：（a）照度、温度和功率；（b）湿度、照度和功率；（c）温度、湿度和功率。这三幅图均将功率变量作为输出。因此，第三幅图显示出较低的相关性，这意味着湿度对最终模型的影响较小。

III. 结果

人工神经网络性能分析考察了包含两个和三个输入变量的情况：(i) 光照、温度和湿度，(ii) 光照和温度，(iii) 光照和湿度，以及 (iv) 湿度和温度；每种组合均采用不同的拓扑结构进行测试。

训练过程采用了 2019 年 1 月 14 日至 2 月 10 日期间的随机数据。表 I 仅展示了根据交叉相关验证中得到的均方根误差筛选出的表现最佳的拓扑结构。表中第一列表示输入变量，第二列表示所用的拓扑结构，第三列和第四列分别表示最大训练周期数和误差水平。最后一列为均方根误差值。请注意，最大训练周期数是根据多次训练实验中的均方根误差表现确定的。

第一种拓扑结构（表I）的输入层包含三个计算元素，隐藏层包含三个元素，并设有一个输出（3:3:1）。该拓扑结构使用随机选取的数据进行了评估，随后使用 Easy-NN 软件导入训练集（700 次测量）、验证集（200 次）和测试集（100 次）。图5a显示了拓扑结构 3:3:1 的估计数据与

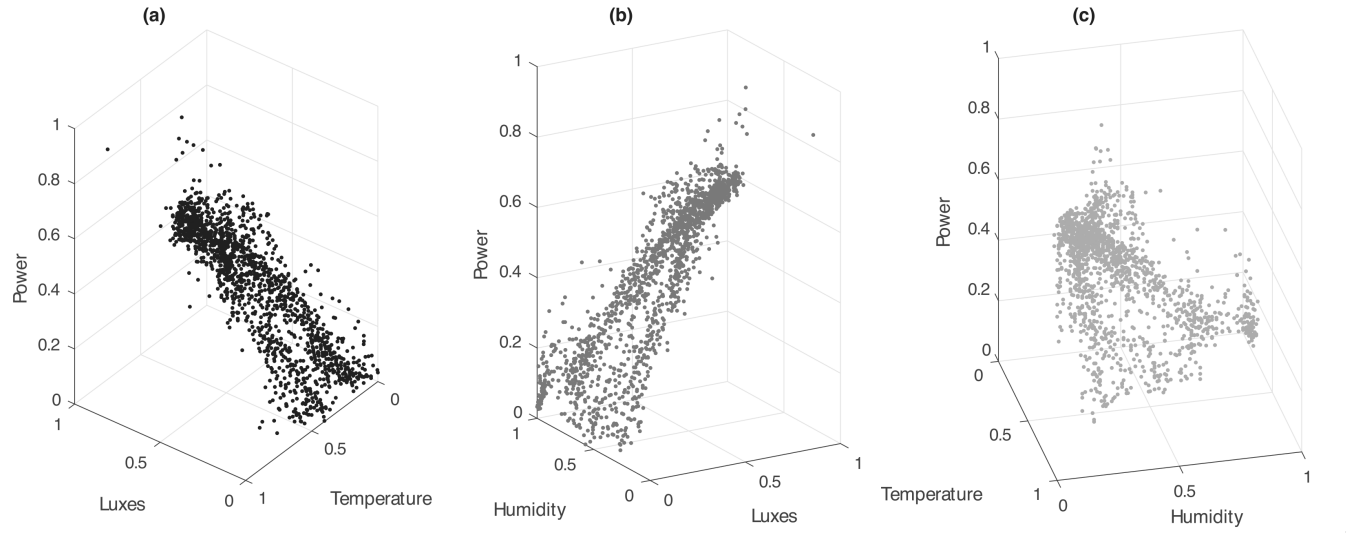


图4：相关变量图：(a) 照度、温度和功率；(b) 湿度、照度和功率；(c) 温度、湿度和功率。

表I：用于预测光伏系统的最佳人工神经网络拓扑结构。

输入变量	ANN 拓扑结构	训练周期	误差水平	RMSE
光照、温度和湿度	3:3:1	5000	0.1	0.255
光照和温度	2:8:1	5000	0.1	0.273
光照和湿度	2:3:1	5000	0.1	0.260
温度和湿度	2:7:1	5000	0.1	1.52

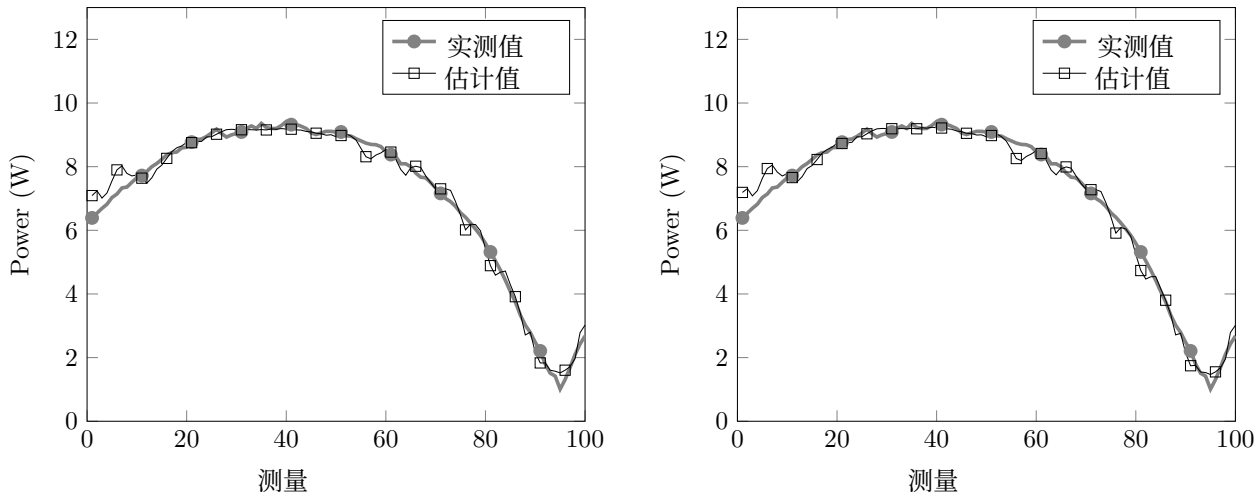
测试数据集的比较。该拓扑结构计算得到的误差为0.255326464。

另一个效果良好的估计器仅使用光照和温度构建，拓扑结构为 2:8:1，误差水平为0.273086254；见图5b。对于基于光照和湿度的估计器，最佳拓扑结构为 2:3:1，误差水平为0.26061261；见图5c。最后，对于使用温度和湿度变量进行的估计，表现最佳的拓扑结构为2:7:1，误差水平为 1.522621379；见图5d。

请注意，效果最佳的神经网络使用了三个变量。不过，在某些数据缺失或损坏时，也可以使用备选神经网络。在本研究中，湿度传感器的问题最多，这是由于温度波动所致；莫雷利亚市尤为明显，导致清晨时段的测量值达到饱和。尽管如此，当所有变量都处于理想状态时，人工神经网络能够给出更准确的估计。

另一方面，研究人员使用包含三个变量的同一数据集开发了多元线性回归器（多元线性回归）。图6对比了人工神经网络、多元线性回归和实际测量值的一日估计结果。请注意，与多元线性回归模型的估计结果相比，人工神经网络的估计结果与光伏系统的实际运行情况更为接近。

根据中心极限定理，在大样本中，无论数据如何，抽样分布都趋于正态分布 [28]。因此，我们进行了方差分析。针对光照、温度和湿度进行多元线性回归分析的方差分析，以确定这些因素的重要性，结果见表II。方差分析的结果得出了置信水平为 95 % 的回归模型。该模型对实际现象的拟合度为 91.77 %。进行这一过程是为了获得具有适当置信水平的学习结果，从而能够适当地预测太阳能电池板的功率水平。



(a) 使用 3 个变量（光照、温度和湿度）进行估计，拓扑结构为 3:3:1

(b) 使用 2 个变量（光照和温度）进行估计 拓扑结构为 2:8:1

(c) 使用 2 个变量（光照和湿度）进行估计，拓扑结构为 2: 3: 1

(d) 使用 2 个变量（温度和湿度）进行估计，拓扑结构为 2:7:1

图5：两种和三种输入变量的最佳拓扑结构下估计数据与实际数据的图形比较

表II：对 x 个自变量和 y 个因变量进行的方差分析。

来源	DF	调整平方和	调整均方	F 值	P 值
回归	3	10159.50	3386.51	5895.20	0.000
光照	1	2761.10	2761.10	4806.50	0.000
温度	1	6.80	6.78	11.81	0.001
湿度	1	47.00	47.01	81.83	0.000
误差	1586	911.10	0.57		
总计	1589	11070.60			

IV. 讨论与结论

在本研究中，我们观察到最合适的神经网络拓扑结构会因.....而变化

由于均方根误差最低，因此选择这些输入变量。通过采用不同的训练周期、输入信息、神经元数量和隐藏层进行实验，以辨别其表现，进而为每组数据选择最合适的配置。所使用的原型存在一些局限，例如测量的分辨率和范围，以及读数流。为验证原型的数值，我们将其与位于墨西哥莫雷利亚技术学院的商业测量站进行了比较。均方根误差为 0.19309，表明原型数据可用于估计。

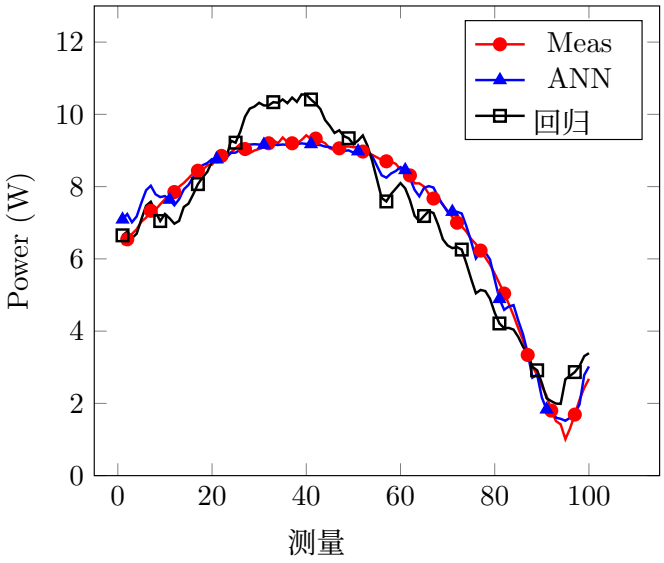


图6：人工神经网络估计与回归模型性能对比图。

数据集的筛选与创建均在 Matlab® 中完成。我们确定了最适合进行即时估算和短期预测的配置，并确定了神经网络最合适的拓扑结构及其参数（计算单元数量、隐藏层数量、输入和输出数量，以及训练周期）。各参数的确定从提出一种随机拓扑结构开始，最终得到 3:3:1 配置，并采用 5,000 个训练周期。该配置可即时了解太阳能电池板在正常条件下的运行情况。本研究结果表明，估计的最佳选择是 3:3:1 神经网络拓扑结构，该网络使用三个变量（光照、温度和湿度），可用于估算太阳能电池板能够产生多少功率。至于预测，最佳配置为 9:4:3:1 神经网络。即使使用两个变量进行估计时的误差比使用三个变量更大，该误差为 0.30320，仍可用于估计。

本文介绍了一种基于人工神经网络（ANN）模型的太阳能预测算法。所提出的模型采用 3:3:1 拓扑结构，训练周期为 5,000 次。研究使用晴空模型和原型的气象数据对神经网络进行训练。研究对原型与莫雷利亚技术学院位于莫雷利亚、墨西哥的气象站进行了比较。均方根误差值证实，利用光照、温度和湿度数据可以进行合理估算。未来的工作将致力于开发更全面的多层人工神经网络模型，考虑降雨因素和一天中的时段，同时使用更大规模的数据集训练人工神经网络模型，以实现更高的预测准确度。此外，还必须提高系统的准确度。

V. 致谢

作者谨向“国家科学与技术委员会”（CONACYT）致谢，感谢其通过资助学生 625015 支持本项目的开展；同时感谢支持项目 6127.17-P 的“墨西哥国家技术学院”（TecNM），以及协助开发电子原型的“国家实验室 SEDEAM”。

VI. 利益冲突

作者声明不存在利益冲突。

参考文献

[1] I. El Kafazi, R. Bannari, A. Abouabdellah, 《能源需求建模与预测》，载于：2016年国际可再生能源与可持续能源会议（IRSEC），IEEE，2016年，第746–750页。[2] Enerdata, 《全球能源统计年鉴》（2019）。网址：https://yearbook.enerdata.net/total-energy/world-consumption-statistics.html [3] K. M. Powell, K. Rashid, K. Ellingwood, J. Tuttle, B. D. Iverson, 《混合式聚光太阳能热发电系统：综述》，《可再生能源与可持续能源评论》80（2017）215–237。[4] P. Jayakumar, 《资源评估手册》，联合国亚洲及太平洋技术转让中心（2009）。[5] E. Muljadi, M. Singh, V. Gevorgian, 《表示光伏发电机的 PSCAD 模块》，技术报告，美国国家可再生能源实验室（NREL），科罗拉多州戈尔登。（2013）。[6] W. C. Tan, L. H. Saw, H. San Thiam, J. Xuan, Z. Cai, M. C. Yew, 《多孔介质/金属泡沫在燃料电池和太阳能发电系统中的应用概述》，《可再生能源与可持续能源评论》96（2018）181–197。[7] Y. Cui, Y. Su, Y. Liu, Y. Liu, D. Smith, 《太阳辐照度与光伏输出变异性指标研究》，载于：2017年IEEE电力与能源学会全体会议，IEEE，2017年，第1–5页。[8] A. Niemi, M. Lehtonen, H. A. R. AbdelHadi, 《太阳辐照度变化作为光伏系统闪变来源的分析》，载于：2017年IEEE欧洲创新智能电网技术会议（ISGT-Europe），IEEE，2017年，第1–6页。

[9] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez- de Pison, F. Antonanzas-Torres, 《光伏功率预测综述》, Solar Energy 136（2016）78–111。[10] A. Mellit, S. A. Kalogirou, 《光伏应用中的人工智能技术：综述》, Progress in Energy and Combustion Science 34（5）（2008）574–632。[11] S. A. Mohammed, S. Shirmohammadi 等, 《基于多模态深度学习的分布式网络延迟测量系统》, IEEE Transactions on Instrumentation and Measurement 69（5）（2020）2487–2494。[12] C. Alippi, A. Ferrero, V. Piuri, 《用于仪器与测量应用的人工智能》, IEEE Instrumentation & Measurement Magazine 1（2）（1998）9–17。[13] D. B. Fogel, 《机器智能》, IEEE Instrumentation & Measurement Magazine 9（3）（2006）12–16。[14] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, Z. Hu, 《面向智能电网能源管理的光伏与太阳能功率预测》, CSEE Journal of Power and Energy Systems 1（4）（2015）38–46。[15] S. J. Russell, P. Norvig, 《人工智能：一种现代方法》, 马来西亚；Pearson Education Limited，2016。[16] E. Alpaydin, 《机器学习：新一代人工智能》, MIT Press，2016。[17] C. M. Bishop 等, 《模式识别神经网络》, Oxford University Press, 1995。[18] M. H. Hassoun 等, 《人工神经网络基础》, MIT Press，1995。[19] C. Monteiro, T. Santos, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado, M. S. Terreros-Olarte, 《基于历史相似性的光伏电站短期功率预测模型》, Energies 6（5）（2013）2624–2643。[20] A. Laouafi, M. Mordjaoui, D. Dib, 《利用人工神经网络预测提前一小时的电力负荷及风光发电》, 载于：2015年可再生能源大会（IREC）, 第六届国际会议，IEEE，第1–6页。[21] M. Brenna, F. Foiadelli, M. Longo, D. Zaninelli, 《利用神经网络预测太阳辐射和负荷功率消耗》, 载于：2017年第六届国际清洁电力会议（ICCEP）, IEEE，2017年，第726–731页。[22] H. Sangrody, M. Sarailoo, N. Zhou, N. Tran, M. Motalleb, E. Foruzan, 《太阳能预测中的天气预报误差》, IET Renewable Power Generation（2017）。[23] T. Verma, A. Tiwana, C. Reddy, V. Arora, P. Devanand, 《通过神经网络与回归模型进行太阳能发电预测的数据分析》, 载于：2016年第七届智能系统、建模与仿真国际会议（ISMS）, IEEE，2016年，第97–100页。[24] Y. Tao, Y. Chen, 《基于遗传算法的神经网络方法进行分布式光伏功率预测》, 载于：2014年国际先进机电系统会议（ICAMechS）, IEEE，2014年，第557–560页。[25] M. J. Sanjari, H. Gooi, 《基于高阶马尔可夫链的光伏发电功率概率预测》, IEEE Transactions on Power Systems 32（4）（2017）2942–2952。[26] V. P. Menon, S. Lokhande, Y. K. Bichpuriya, 《随着太阳能光伏渗透率提高，对公用事业需求预测进行修正》, 载于：2017年IEEE亚洲创新智能电网技术会议（ISGT-Asia）, IEEE，2017年，第1–6页。[27] J. Corona-Ventura, 《利用物联网（IoT）表征太阳能电池板》, Morelia Institute of Technology，2019。[28] A. Ghasemi, S. Zahediasl, 《统计分析中的正态性检验：面向非统计学家的指南》, International Journal of Endocrinology and Metabolism 10（2）（2012）486。