

杨迪,刘林青,马红明,等. 基于特征聚类的光伏集群出力时空随机模拟方法[J]. 湖南科技大学学报(自然科学版), 2023, 38(1):76-84. doi:10.13582/j.cnki.1672-9102.2023.01.010

YANG D, LIU L Q, MA H M, et al. A Spatiotemporal Stochastic Simulation Method for Photovoltaic Cluster Based on Feature Clustering [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2023, 38(1):76-84. doi:10.13582/j.cnki.1672-9102.2023.01.010

基于特征聚类的光伏集群 出力时空随机模拟方法

杨迪^{1*}, 刘林青¹, 马红明¹, 王立斌¹, 马浩¹, 冀明¹, 张悦仙²

(1. 国网河北省电力有限公司营销服务中心, 河北 石家庄 050000;

2. 西安交通大学 智能网络与网络安全教育部重点实验室, 自动化科学与工程学院, 陕西 西安 710049)

摘要: 随着分布式光伏集群的建设和集群控制的需求, 不仅需要单个光伏电站的信息, 而且需要光伏集群内各站点的综合信息, 亟须发展光伏集群出力时空的随机模拟模型. 而光伏集群内各站点出力在时间和空间上的相关性, 使得光伏集群出力特性的随机模拟模型的准确性不能保证. 为此, 文章提出了一种基于特征聚类的光伏集群出力时空随机模拟模型. 该方法基于 k -means 特征聚类方法, 考虑不同站点出力时空相关性特征, 将光伏出力以天气类型划分为 4 类, 并依此构建单站点出力的马尔可夫链的时序模型, 利用不同站点间时空的相似性完成对光伏集群的数据模拟, 为分布式光伏集群的建设和集群控制提供数据参考. 基于河北电网部分实际光伏电站的仿真计算, 验证了所提数据模拟方法的正确性和有效性.

关键词: 光伏集群; 时空相关性; 聚类分析; 马尔可夫链

中图分类号: TM734

文献标志码: A

文章编号: 1672-9102(2023)01-0076-09

A Spatiotemporal Stochastic Simulation Method for Photovoltaic Cluster Based on Feature Clustering

YANG Di¹, LIU Linqing¹, MA Hongming¹, WANG Libin¹, MA Hao¹, JI Ming¹, ZHANG Yuexian²

(1. State Grid Hebei Electric Power Company Marketing Service Center, Shijiazhuang 050000, China;

2. Ministry of Education Key Laboratory of Intelligent Networks and Network Security, School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: With the construction of distributed PV clusters and the demand of cluster control, not only the information of individual PV sites but also the comprehensive information of each site within the PV cluster is needed, and there is an urgent need to develop the stochastic simulation model of PV cluster power output in time and space. Within the PV cluster, the temporal and spatial correlation of the power output in each site makes the accuracy of the stochastic simulation model of PV cluster output characteristics cannot be guaranteed. To this end, a feature clustering-based stochastic simulation model of PV cluster power output in time and space is proposed in this paper. The method is based on k-means feature clustering method, which considers the spatio-temporal correlation characteristics of the power output of different sites. It divides the PV power output into four

收稿日期: 2021-07-07

基金项目: 国网河北省电力有限公司科技项目资助(SGHEDK00DYJS1900304); 国家自然科学基金资助青年项目(61903293)

* 通信作者, E-mail: 654703101@qq.com

categories based on the weather type, and constructs the Markov chain temporal model of the power output of a single site accordingly, and completes the data simulation of the PV cluster using the spatio-temporal similarity between different sites to provide data reference for the construction and cluster control of distributed PV clusters. Based on the simulation calculation of some actual PV plants in Hebei power grid, the correctness and effectiveness of the proposed data simulation method are verified.

Keywords: photovoltaic clusters; spatio-temporal correlation; cluster analysis; Markov chains

长时间以来,电力系统发展都伴随着对化石能源等不可再生能源的开采和利用,但是不可再生能源储量有限,且燃烧时会释放有害气体,造成环境破坏、温室效应等问题,因此,提升可再生能源利用的能力,实现可持续发展就成为了解决这一问题的关键途径。目前,太阳能、风能等可再生能源的应用前景广阔,是低碳电网建设的关键所在^[1-2]。在电力产业里,光伏发电以其高效、清洁等优点而被广泛运用于发电系统中。因此,光伏电站输出功率的特性对电力系统运行和控制有着至关重要的影响。

光伏发电是一种利用光伏电池的光生伏效应,将太阳辐射的光能转变为电能的技术^[3]。目前,光伏发电技术常利用聚光器的高精度自动跟踪系统来提升热接受率^[4],利用大容量电池储能系统来增加系统的稳定性^[5],但由于四季更替、多变天气等外界环境因素的影响,光伏数据通常具有比较强的随机性和间歇性,无疑会为光伏系统的控制和运行带来挑战^[6]。考虑到光伏出力特性^[7],充分挖掘光伏数据价值,实现其优化配置对于提高电力系统运行能力有着极大的意义。

针对光伏数据的特性研究都需要大量的数据,而数据样本量不足就成了一个比较大的限制,所以必须对光伏数据进行模拟和建模,扩充光伏数据的样本容量,供后期搭建电力数据智能分析平台使用。目前研究光伏数据模拟的方法主要分为 2 种。第一类方法基于太阳辐照度进行建模和模拟,比较流行的有正态分布模型和 beta 分布模型。文献[8]在此基础上利用条件核密度估计方法,建立一种非参数概率模型;文献[9]考虑了光伏电站地理位置、辐射度、设备型号、光伏组件温度特性、表面遮挡程度等因素,利用不同倾角和朝向利用系数,对于光伏电站整体发电效率进行模拟。但是光伏系统的能量转化受到多种因素的影响,此种方法难以综合考虑全部因素,通常所得到的结果并不准确。第二类方法则是基于历史光伏出力数据,利用统计方法,采用数学拟合回归等技术手段生成模拟序列,这种方法可以省去能量转化过程,简化算法程序,同时提高算法的精确程度。文献[10]利用马尔可夫链模型,综合光伏数据天气特性进行模拟,反映出了光伏出力的季节特性和日特性;文献[11]在此基础上提出利用模糊 C 均值聚类进行场景划分的方法,克服抽样过程中较强的随机性;文献[12]通过分析光伏电池的物理模型来确定要收集的气象特征,采用神经网络对各个类别进行训练。但是,这 3 篇文章都将光伏出力的功率作为构成马尔可夫状态的唯一指标,当天气状况转变比较复杂时,这样的区分方法会比较难刻画出数据 1 d 的变化情况,同时,目前针对光伏数据的模拟研究主要是建立在单站点基础上,没有考虑多站点由于地理和空间相似性所带来的联系,可以用于简化多站点情况下的数据模拟。

单个光伏电站数据模拟通常基于单个站点的历史数据,所模拟出来的数据也是贴合单个站点的。当需要模拟的站点数量较多,同时站点在地理和空间上相近时,那么仅采用单站点的模拟方法就忽视了多个站点之间的相似性。光伏集群的数据模拟方法则建立在对多个站点同时进行数据模拟的思想,先利用各个站点的历史数据进行站点聚类,将光伏出力相似度高的站点聚类在一起,再对同簇内站点同时进行数据模拟。例如,在对某个省在省内各个地区建立的多个光伏站点进行数据模拟时,若采用单站点模拟的方法对所有站点进行模拟,计算量无疑是巨大的。但是,由于处在同一省份,各个站点之间存在地域上的相近,光伏出力数据也呈现出一定的相似性。可以利用光伏集群的数据模拟方法,将出力数据相似性高的光伏站点统一处理,进行数据模拟,减小计算量,并且将同簇内多个站点的数据同时作为历史数据进行模拟,可以提高模拟数据的合理性,不会因站点某一日出现极端或不合理的光伏出力数据导致最终模拟数据出现比较大的波动。

本文提出一种针对多电站的光伏数据模拟方法,在对单站点数据模拟的基础上,利用相近站点间空间和距离的相似性完成对站点的数据模拟。此方法基于特征聚类和马尔可夫转移概率矩阵,使用均值、标准差、偏度、锋度以及波动程度多个维度来确定某日光伏数据的马尔可夫状态,同时使用 k -means 法将多站

点数据按类区分,综合同类站点的历史数据进行建模,从而得到同类站点的模拟数据.所提出的方法能够简化多站点的模拟数据计算流程,降低计算量,将站点聚类后簇内各站点的历史数据共同用来模拟新数据,解决单个站点数据过少,模拟不够合理的问题.

1 光伏数据日特征分析

影响光伏发电功率的气象因素众多,但最主要的因素是太阳辐射度^[13],其周期性和随机性相结合的特点使得光伏数据呈现出明显的规律性和波动性.规律性是由于日地运动引起的,主要表现为光伏数据在上午时段随时间推移呈现上升趋势,而在下午时段则呈现下降趋势,在无光照情形下光伏发电功率为0.波动性则是由于云层移动、阵雨等微尺度气象变化引起的^[14],一般体现在光伏数据曲线并非平稳过渡.在几种典型天气状况下,光伏数据表现出较大的差异,导致光伏发电适应性的不同.图1和图2分别给出2个地理位置相近的站点在2019年不同天数下的光伏数据曲线,通过对比两站点数据可以得出:

1) 不同天气类型下光伏数据表现出的差异较大,4条曲线的区分度高,可以通过聚类的方法将站点整年数据进行区分,每日的数据对应一种天气类型,根据平均值由大到小可分为晴、阴、小雨、大雨.

2) 距离相近站点的光伏数据具有相似特性,说明在相同地理位置的天气状况变化规律一致,可以利用聚类方法将站点数据变化趋势相近的站点作为同一类站点来处理.

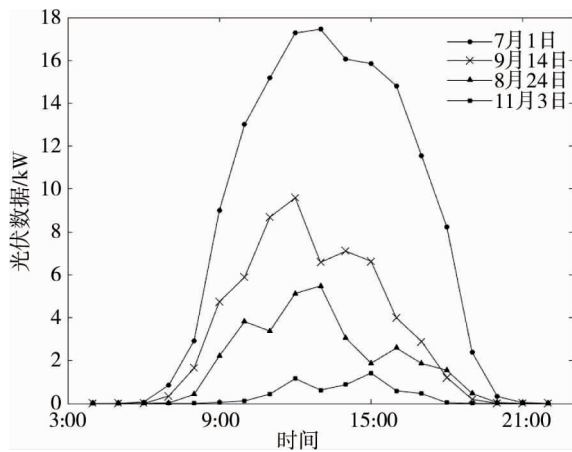


图1 A 站点在不同日的光伏数据表现

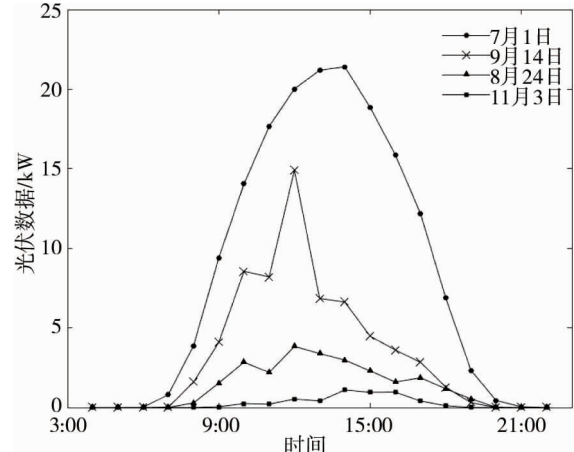


图2 B 站点在不同日的光伏数据表现

本文选用的光伏数据以1h为间隔,由于夜晚光伏发电功率过低,故去除每日21:00—3:00的数据,从每日0点到次日0点共记录19个数据,对于这样一个时间序列,如果直接将原始数据集作为日特征,那么其向量维数显然是巨大的,无法直接作为聚类等算法的输入,因此有必要对光伏数据进行特征提取.定义第*i*日光伏数据日特征向量 F_i 为

$$F_i = \{f_1, f_2, f_3, f_4, f_5\}. \quad (1)$$

式中: f_1 为每日光伏数据的平均值,用以反映数据的集中趋势,公式描述为

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x(i); \quad (2)$$

f_2 为标准差,可以量化时间序列的离散程度,公式描述为

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x(i) - \bar{x})^2}; \quad (3)$$

f_3 为偏度,是一个时间序列的概率分布关于其平均值的不对称性的度量,公式描述为

$$\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{x(i) - \bar{x}}{\sigma} \right)^3; \quad (4)$$

f_4 为峰度,用以表征概率密度分布曲线在平均值处峰值的高低,公式描述为

$$\beta = \frac{1}{N} \sum_{i=1}^N \left(\frac{x(i) - \bar{x}}{\sigma} \right)^4 - 3; \quad (5)$$

f_5 为移动波动强度 (MFI, Moving Fluctuation Intensity), 用以表征光伏数据的波动强度, 此处引入波动评价指标平均波动幅度 (AFM, Average Fluctuation Magnitude) 和反向波动计数 (RFC, Reverse Fluctuation Count) 以衡量数据的波动指标^[15]. 其中 AFM 定义:

$$\text{AFM} = \frac{1}{N} \sum_{i=1}^N \text{abs}[x(i+1) - x(i)]; \quad (6)$$

RFC 定义: 当 $[x(i+1) - x(i)][x(i) - x(i-1)] < 0$ 时, RFC 的计数值加 1. 由此得到了可以表征 1 d 光伏数据的五维日特征向量, 显著减小聚类算法的计算量.

2 光伏集群出力时空模拟

根据分布式光伏集群建设和集群控制的需求, 针对光伏数据模拟的研究不仅需要单个光伏站点的信息, 还需要光伏集群内各站点的综合信息. 而目前针对光伏数据模拟的方法主要集中于单站点, 即忽视多个站点之间的相关性而仅基于单站点的历史数据做出模拟. 因此, 本文提出一种基于特征聚类的光伏集群出力时空模拟算法, 利用 2 次聚类方法, 可以同时考虑站点的地理环境相似性和天气类型, 从而利用马尔可夫链的形式描述光伏集群出力的随机特性, 进而实现对光伏集群的数据模拟.

2.1 站点环境聚类与天气类型聚类

由于太阳辐照度与纬度位置、天气状况、海拔高低、日照长短等因素直接相关, 所以站点所在地域和地理环境对光伏数据有直接的影响. 当针对多站点进行研究时, 应当考虑不同站点间的地理环境相似性, 利用光伏数据将环境相似性高的站点聚类研究, 最终对于每个簇给出模拟数据, 可以有效减小模拟数据计算量. 同时, 还可以起到扩充历史数据容量的作用, 单个站点的历史数据量比较小时, 利用簇内所有站点历史数据模拟出来的数据会更具有可信度, 本文利用 k -means 聚类方法, 根据各站点的历史数据对站点按照地理环境分类, 规定聚类个数为 4.

定义站点 j 的所有日平均特征向量: $\mathbf{F}^j = \{\bar{f}_1, \bar{f}_2, \bar{f}_3, \bar{f}_4, \bar{f}_5\}$, 利用所有站点的日平均特征向量的平均值对站点进行聚类, 将 m 个站点划分为 4 个地域簇 $\mathbf{S} = \{s_1, s_2, s_3, s_4, s_5\}$, 并使得集合中方差最小.

将簇内站点的光伏数据以天气类型分为 4 种: 晴、阴、小雨和大雨, 分别对应图 1 中的 4 条曲线. 晴天的光伏发电适宜性最好, 对应数据值最高的一条曲线; 大雨天气的光伏发电适宜性最差, 对应数据值最低的一条曲线. 为了确定某日光伏数据对应的天气状况, 本文采用了 k -means 聚类方法, 规定聚类个数为 4.

k -means 聚类是一种统计技术, 它将聚类中的对象进行分组, 并且属于同一个聚类的对象尽可能相似. 根据给定的第 i 日观测光伏数据的归一化五维特征向量 $\mathbf{F}_i = \{f_1, f_2, f_3, f_4, f_5\}$, 对所有的天数 n 进行聚类, 将 n 组数据划分为 4 个簇 $\mathbf{C} = \{c_1, c_2, c_3, c_4, c_5\}$, 分别对应晴、阴、小雨和大雨, 并使得集合中方差最小, 具体实现算法如下所示:

第一步: 将数据集 $\{F\}$ 的 k 个随机点设置为初始的聚类中心 $\text{center}(j)$, $j = 1, 2, 3, 4$;

第二步: 计算每个点与每个簇中心点之间的欧几里得距离 $d(i, j)$, $i = 1, \dots, n, j = 1, 2, 3, 4$;

第三步: 对于任意一点 \mathbf{F}_i , 将其划分到距离它最近的簇中心 $\text{center}(j)$ 所在簇 $c(j)$, 即当 $d(i, j) = \min\{d(i, j), j = 1, 2, \dots, k\}$ 时, 点 \mathbf{F}_i 会被分到第 j 个簇;

第四步: 对所有的点 j 遍历结束后, 将每个簇中的所有点的平均值设为新的簇中心;

第五步: 返回到第二步, 直到结束.

2.2 马尔可夫链与转移概率矩阵

分布式光伏的出力特性具有典型的概率分布特点^[16]. 由于太阳光的辐射具有随机性, 因此, 与其相关的光伏数据可以看作是随机过程. 本文将光伏数据的 4 种发电适宜性状态视为马尔可夫链. 马尔可夫链是一种具有离散时间和离散状态的马尔可夫随机过程, 下一时刻的系统状态只与当前时刻的状态有关, 而与以前的状态无关. 定义离散时间序列 $X(i)$, $i = 1, \dots, n$ 的离散状态空间为 $E = \{e_1, e_2, \dots, e_k\}$, $X(t) = e$ 表示在时刻 t 随机过程保持在状态 e .

定义 $P(i, j)$ 为从 $X(t) = i$ 到 $X(t) = j$ 的转移概率, 假设整体数据集有 n 天, 且由状态 i 转移到 j 这件事情发生了 m 次, 那么 $P(i, j)$ 的计算公式为

$$P(i,j) = P\{X(t+1) = j | X(t) = i\} = \frac{m}{n}. \quad (7)$$

计算所有状态的转移概率就可以得到马尔可夫转移概率矩阵:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (8)$$

该矩阵在状态转移的过程中保持不变,且各行之和为1.

2.3 光伏集群出力生成算法

现在可以根据转移概率矩阵和原始光伏数据生成连续的模拟数据,算法如下:

第一步:随机选择初始状态 e_1 , 设置模拟的天数 n_e , 生成数据集 $S_d = \{e_1\}$, 计数器值为 Counter = 1;

第二步:在 $0 \sim 1$ 之间取一个随机数 r , 找到满足 $\sum_{i=1}^{e_2-1} P(e_1, e) < r < \sum_{i=1}^{e_2} P(e_1, e)$ 的第2天状态 e_2 ;

第三步:将第2天状态 e_2 加入数据集 S_d 中,并重新将 e_2 作为初始状态,将计数器值改变,使得 Counter = Counter+1;

第四步:检查 Counter 值,当 Counter > n_e 时,转到第五步,否则返回第二步;

第五步:对于 S_d 中每个状态元素,随机在原始光伏数据中选择对应于该状态的一天作为模拟数据;

第六步:将产生的 n_e 个数据集连接在一起得到要求的模拟数据.

整体光伏集群的数据模拟方法流程图如图3所示.

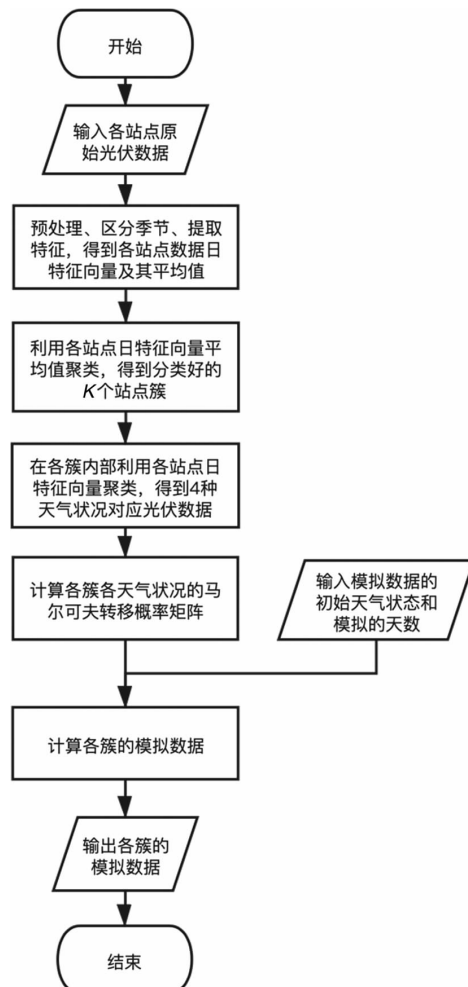


图3 光伏集群的数据模拟方法

3 仿真实例分析

3.1 光伏数据处理规模

本文选取河北 72 个电站的历史光伏数据作为原始数据,对每个站点 7 月 1 日—12 月 30 日 183 d 的光伏发电功率进行分析,共取得了 13 176 d 的数据,该数据的采样时间间隔为 1 h.注意到每日 21:00—3:00 的光伏发电功率过低,无法满足发电所需,因此删除了此时间段的数据,每日保留 19 个采样点的值.同时由于各季节的光照呈现出显著的差异,为了体现光伏数据的季节特性,按照月份分为 2 个季度,认为 7 月—9 月为夏季,10 月—12 月为冬季.

3.2 k -means 聚类

本文中利用 k 均值算法分别对站点和天气类型进行了聚类.下面以河北 72 个光伏电站的实测数据为例进行分析.

针对站点地域环境进行聚类时,首先提取单个站点一日的 19 个时刻值用于计算 1.1 中提到的 5 个特征值,将此站点 183 d 的特征值平均值作为聚类参数,选取聚类个数 $k=4$,得到 4 个簇中包含站点数量分别为 37,5,17,13,给 4 个簇的站点分别命名为 Cluster I, Cluster II, Cluster III, Cluster IV,后期针对数据的处理都在同簇内进行.

在同一个簇内,针对天气类型进行聚类.将数据分为夏、冬两季后,利用每日的 5 个特征值,对同簇同季节数据进行 k 值为 4 的聚类,得到的 4 个簇分别对应 4 种天气状况:晴、阴、小雨、大雨.为直观显示聚类的结果,计算 4 类站点 2 个季节同一天气状况下的所有天数的平均光伏数据,如图 4 和图 5 所示.

图 4 和图 5 分别展示了夏季和冬季,4 个地域簇中在同一天气下的所有天数光伏数据平均值.以图 4 中的 Cluster I 和 Cluster II 为例,2 幅图片的幅值差异比较大,Cluster II 中晴天的光伏数据平均值幅值在正午时间能够达到 25 kW 以上,而 Cluster I 的晴天的光伏数据平均值幅值在正午时间也仅能在 4 kW 以下.这说明 Cluster II 中所包含的站点相较于 Cluster I 中所包含的站点可能海拔更高,空气更稀薄,大气透明度较高,对太阳辐射的削弱作用小,可以接收更多的太阳光照.将 4 个簇的光伏数据简单比较可知:光照资源最充足的地域为 Cluster II 中站点所在位置,光照资源最缺乏的地域为 Cluster I 中站点所在位置.

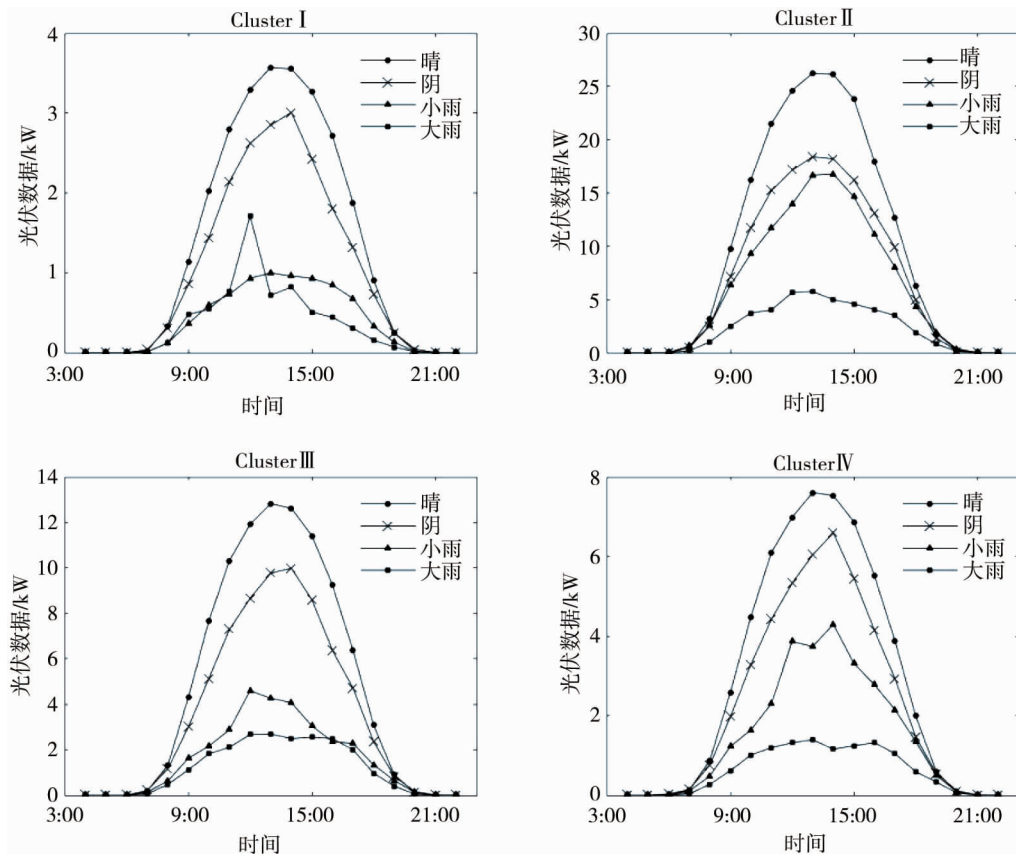


图 4 夏季各簇同一天气所有天数光伏数据平均值

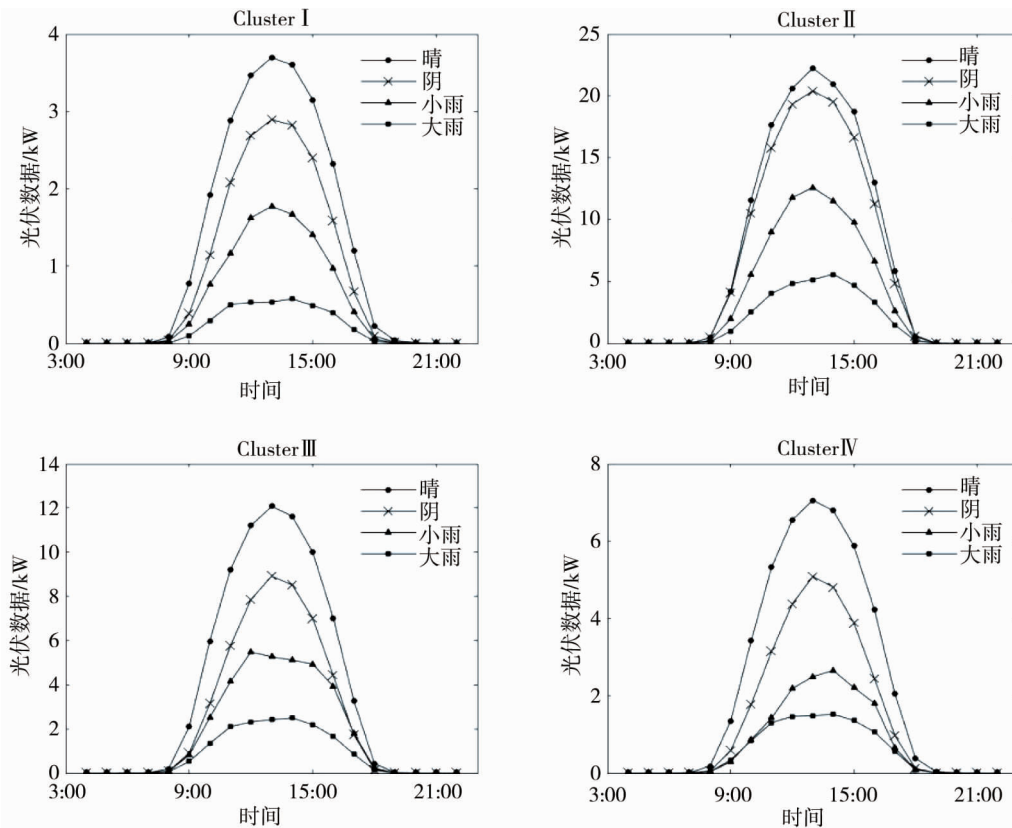


图5 冬季各簇同一天气所有天数光伏数据平均值

将图4和图5中4个簇中曲线幅值对比可见:各簇光伏数据幅值差距较大,说明聚类算法可以很好地将不同类站点进行区分.同时簇内4种天气状况对应的曲线区分比较清晰,分别对应4种光伏发电的适宜性状态,将各线条对应天数总结在表1中,用于后续计算.

表1 各簇各季节天气状况数量

天气	Cluster I		Cluster II		Cluster III		Cluster IV	
	夏	冬	夏	冬	夏	冬	夏	冬
晴	679	412	105	101	293	324	1596	1 033
阴	332	415	136	108	293	574	996	990
小雨	85	148	54	47	242	189	718	961
大雨	192	299	73	108	736	460	94	383
总计	1 288	1 274	368	364	1 564	1 547	3 404	3 367

3.3 马尔可夫转移概率矩阵

为了克服单个站点数据量过少,导致统计特性不明显的问题,综合同簇站点所有数据来计算马尔可夫转移概率.利用式(7)计算天气状况变化的概率,其中的*i*和*j*分别代表2种天气状况,表2展示了各簇站点综合夏、冬两季的马尔可夫转移概率矩阵.

3.4 仿真结果分析

利用表2中的数据及光伏数据的生成算法,分别得到模拟数据的夏、冬两部分,进行拼接后则为此簇光伏数据整体的模拟数据.模拟数据生成算法的核心思想是:将一簇数据按照天气状况分为4类后,利用马尔可夫转移概率矩阵确定天气状况之间的连续变化情况,再利用随机数和累计状态转移概率矩阵计算出模拟数据的天气类型,最后基于当日模拟数据对应的天气类型,在历史数据中随机挑选相应类型填入模拟数据中,完成从天气类型到光伏数据的转变.在此转变过程中,为了保证光伏数据符合时间变化规律,通常会在相应天气类型下选择时间最靠近模拟日的数据,例如,当7月28日的模拟天气类型为晴时,会在历史数据中搜索天气为晴同时时间上最靠近7月28日的光伏数据作为模拟数据.

为了清晰展示模拟数据的效果,选取 Cluster III 的模拟数据和簇内各站点的历史数据,计算各月数据平均值绘制在图 6 中.可以发现模拟数据整体趋势走向符合簇内历史数据,基本呈现出了从夏季到冬季的下降趋势,符合季节规律,在各月份的转折处也表现出和历史数据相同的转折趋势.将模拟数据和历史数据的特征值记录在表 3 中,观察可见,模拟数据的特征值在夏季、冬季和全年 3 个角度都同历史数据的吻合度较好,在还原站点特征的基础上还保留了自己的差异性,得到的模拟效果较好.

表 2 各簇各季节的天气状况变化对应的马尔可夫转移概率矩阵

地域簇	夏	冬
Cluster I	$P = \begin{bmatrix} 0.327 & 0.154 & 0.421 & 0.097 \\ 0.208 & 0.317 & 0.343 & 0.130 \\ 0.200 & 0.118 & 0.644 & 0.036 \\ 0.588 & 0.011 & 0.352 & 0.041 \end{bmatrix}$	$P = \begin{bmatrix} 0.609 & 0.181 & 0.164 & 0.044 \\ 0.192 & 0.503 & 0.214 & 0.090 \\ 0.227 & 0.274 & 0.277 & 0.220 \\ 0.101 & 0.297 & 0.418 & 0.182 \end{bmatrix}$
Cluster II	$P = \begin{bmatrix} 0.242 & 0.198 & 0.157 & 0.371 \\ 0.120 & 0.615 & 0.085 & 0.178 \\ 0.163 & 0.119 & 0.334 & 0.382 \\ 0.129 & 0.034 & 0.181 & 0.654 \end{bmatrix}$	$P = \begin{bmatrix} 0.448 & 0.140 & 0.289 & 0.121 \\ 0.212 & 0.545 & 0.191 & 0.050 \\ 0.203 & 0.213 & 0.425 & 0.157 \\ 0.369 & 0.195 & 0.260 & 0.173 \end{bmatrix}$
Cluster III	$P = \begin{bmatrix} 0.272 & 0.198 & 0.157 & 0.371 \\ 0.120 & 0.615 & 0.085 & 0.178 \\ 0.163 & 0.119 & 0.334 & 0.382 \\ 0.129 & 0.034 & 0.181 & 0.654 \end{bmatrix}$	$P = \begin{bmatrix} 0.239 & 0.372 & 0.271 & 0.117 \\ 0.154 & 0.397 & 0.282 & 0.165 \\ 0.086 & 0.261 & 0.562 & 0.089 \\ 0.069 & 0.187 & 0.193 & 0.549 \end{bmatrix}$
Cluster IV	$P = \begin{bmatrix} 0.580 & 0.127 & 0.003 & 0.288 \\ 0.289 & 0.365 & 0.102 & 0.243 \\ 0.255 & 0.446 & 0.042 & 0.255 \\ 0.415 & 0.215 & 0.011 & 0.347 \end{bmatrix}$	$P = \begin{bmatrix} 0.322 & 0.230 & 0.188 & 0.249 \\ 0.283 & 0.433 & 0.220 & 0.063 \\ 0.200 & 0.196 & 0.574 & 0.045 \\ 0.462 & 0.295 & 0.148 & 0.094 \end{bmatrix}$

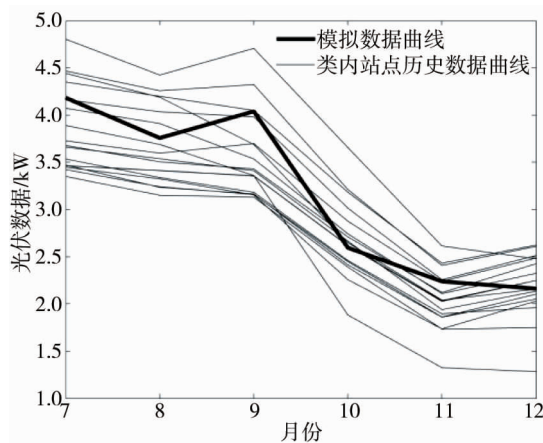


图 6 Cluster III 的月度平均曲线

表 3 Cluster III 的统计特性比对

Cluster III	平均值	标准差	偏度	峰度	MFI	最大值
夏季原始数据	3.698 9	4.073 0	0.680 0	2.162 0	83.001 4	10.932 2
夏季模拟数据	3.993 3	4.398 1	0.637 6	1.981 6	83.463 7	11.689 4
冬季原始数据	2.314 4	3.068 5	0.976 8	2.555 5	51.380 2	8.282 1
冬季模拟数据	2.334 1	3.119 6	0.945 9	2.397 9	55.330 3	8.490 3
全年原始数据	3.010 4	3.573 5	0.827 6	2.357 7	67.277 2	9.614 4
全年模拟数据	3.168 2	3.762 3	0.790 9	2.188 6	69.473 8	10.098 6

5 结论

1) 提出了一种对多电站的基于特征聚类方法和马尔可夫转移概率矩阵的光伏数据模拟方法,通过统计光伏电站历史数据进行建模,对大量的光伏功率数据进行模拟,用于扩充样本容量和对缺失数据的填充.

2) 最终得到的模拟数据特征值在夏季、冬季和全年3个角度都同历史数据的吻合度较好,在还原站点特征的基础上还保留了自己的差异性,得到的模拟效果较好,结果表明本文提出的多站点数据模拟方法得到了较好的效果.

参考文献:

- [1] 孙建梅,李龙龙.低碳电网建设研究现状及未来发展趋势[J].电网与清洁能源,2019,35(3):21-30.
- [2] 姜曼,杨司玥,刘定宜,等.中国各省可再生能源电力消纳量对碳排放的影响[J].电网与清洁能源,2020,36(7):87-95.
- [3] 上官小英,常海青,梅华强.太阳能发电技术及其发展趋势和展望[J].能源与节能,2019(3):60-63.
- [4] 彭长清,彭佑多,胡燕平,等.大功率单碟式太阳能聚光器跟踪控制系统研究及实现[J].湖南科技大学学报(自然科学版),2012,27(3):28-34.
- [5] 程孟增,梁毅,宋颖巍,等.大容量电池储能系统孤网运行控制策略[J].湖南科技大学学报(自然科学版),2015,30(3):76-82.
- [6] 赖昌伟,黎静华,陈博,等.光伏发电出力预测技术研究综述[J].电工技术学报,2019,34(6):1201-1217.
- [7] 刘文霞,何向刚,吴方权,等.新能源发电出力特性指标及其数据化应用[J].电网与清洁能源,2020,36(9):85-92.
- [8] 金黎明,周宁,冯丽,等.光伏发电系统的时序概率模型研究[J].电力建设,2016,37(7):27-32.
- [9] 柯学进.光伏电站影响因素及发电量分析验证方法的探讨[J].电力设备管理,2021(4):125-127.
- [10] 丁明,鲍玉莹,毕锐.应用改进马尔可夫链的光伏出力时间序列模拟[J].电网技术,2016,40(2):459-464.
- [11] 江雪辰,朱俊澎,袁越,等.基于新型场景划分与考虑时序相关性的光伏出力时间序列模拟方法[J].电力建设,2018,39(10):63-70.
- [12] ZHU F, YANG Q, YAN W. Photovoltaic Power Generation Prediction Using Data Clustering and Parameter Optimization [C]// 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2019: 1-5.
- [13] 吕耀堂. 基于数据挖掘的光伏发电建模及功率预测研究和应用 [D]. 广州:华南理工大学,2017.
- [14] 赵亮,黎嘉明,艾小猛,等.光伏出力随机性分量的提取和统计特征分析[J].电力系统自动化,2017,41(1):48-56.
- [15] ZHANG W, LIU Z. Simulation and analysis of the power output fluctuation if photovoltaic modules based on NREL one-minute irradiance data[C]// International Conference on Materials for Renewable Energy & Environment, 2014, 1:21-25.
- [16] 王洪坤,葛磊蛟,李宏伟,等.分布式光伏发电的特性分析与预测方法综述[J].电力建设,2017,38(7):1-9.