

# 数据驱动的电力系统运行方式分析

侯庆春<sup>1</sup>, 杜尔顺<sup>1\*</sup>, 田旭<sup>2</sup>, 刘飞<sup>2</sup>, 张宁<sup>1</sup>, 康重庆<sup>1</sup>

- (1. 清华大学电力系统及发电设备控制和仿真国家实验室, 北京市海淀区 100084;
2. 国网青海省电力公司经济技术研究院, 青海省 西宁市 810008)

## Data-driven Power System Operation Mode Analysis

HOU Qingchun<sup>1</sup>, DU Ershun<sup>1\*</sup>, TIAN Xu<sup>2</sup>, LIU Fei<sup>2</sup>, ZHANG Ning<sup>1</sup>, KANG Chongqing<sup>1</sup>

- (1. State Key Laboratory of Control and Simulation of Power System and Generation Equipments, Tsinghua University, Haidian District, Beijing 100084, China; 2. State Grid economic and Technological Research Institute of Qinghai Electric Power Company, Xining 810008, Qinghai Province, China)

**ABSTRACT:** The power system operation mode changes substantially due to increasing renewable energy penetration and power electronic devices integration in power system. New operation modes are emerging in power system operation, planning, protection, and stability analysis, which can hardly be handled by traditional model and experience-driven method. To this end, this paper proposed using data-driven method to analyze power system operation mode and its variation based on high dimensional simulated chronological power system operation data. This method first preprocesses the data, identifies representative operation mode patterns and its number using clustering algorithm, and then extracts key features using t-SNE for operation modes visualization. In addition, several indices were proposed to evaluate power system operation modes space dispersion, seasonal consistency, and time variation. A case study on Gansu provincial power system in China showed the impacts of high renewable energy penetration on power system and validated the proposed data-driven method. The results show that the space dispersion and pattern number of operation modes are significantly increasing as renewable energy penetration increases. In addition, the operation mode are less correlated with season under high renewable energy penetration.

**KEY WORDS:** data-driven; high renewable energy penetration; power system operation mode; power system simulation; dimension reduction; clustering

**摘要:** 随着电网中高比例可再生能源及电力电子设备渗透率的提高, 电力系统的运行形态将发生深刻变化, 基于经验选择的电力系统方式分析, 越来越难以应对电力系统运行、规划、保护和稳定分析的要求。为此, 该文基于电力系统时序

运行模拟数据, 提出了数据驱动的电力系统运行方式及其形态变化分析方法。该方法首先对高维运行模拟数据进行预处理, 通过聚类和紧密度指标饱和点分别辨识典型运行模式及其数量, 利用 t-SNE 降维算法提取主要特征并可视化运行方式的时空变化规律; 接着, 从定量的角度, 提出了描述电力系统运行方式分散性、季节一致性、时序多变性的评估指标。最后, 以甘肃电网实际算例系统分析了高比例可再生能源对电力系统运行方式的影响, 验证了文中数据驱动方法的有效性。实验结果表明, 电力系统运行方式的分散性及典型运行方式的数量会随着可再生能源渗透率的上升而迅速增加。此外, 电力系统运行方式与季节的耦合性逐渐减弱。

**关键词:** 数据驱动; 高比例可再生能源; 电力系统运行方式; 电力系统运行模拟; 降维; 聚类

## 0 引言

大力发展可再生能源成为构建新一代低碳化、清洁化电力系统重要趋势<sup>[1-2]</sup>。但高比例可再生能源的强不确定性及电力电子设备的大量接入给电力系统带来了多重挑战, 例如运行方式多样化、电网潮流双向化、稳定机理复杂化等<sup>[3-4]</sup>。此外, 可再生能源消纳的压力及日益增长的调峰调频需求使储能等灵活性资源在电力系统中的作用逐渐凸显<sup>[5-7]</sup>。这些新的因素开始主导电网的运行方式, 传统基于模型和经验的电力系统分析方法越来越难以应对电力系统运行、规划、保护和稳定分析中海量多变的运行方式<sup>[8-9]</sup>。由于传统电力系统的运行方式主要受负荷或季节性水电出力的影响, 具有相对固定的模式, 在电力系统规划中往往利用“冬大冬小, 夏大夏小, 丰水枯水”选取典型运行方式进行经济性、安全性和稳定性评估。但在高比例可再生能源和电力电子设备接入下, 按传统方式选取的电力系统典

基金项目: 国家重点研发计划项目(2016YFB0900100); 国家自然科学基金项目(51907100, 51677096)。

National Key R&D Program of China (2016YFB0900100); National Natural Science Foundation of China (51907100, 51677096)。

型运行场景难以还原真实的情况。因此,如何在运行和规划中确定典型运行方式及其数量?如何在保护和稳定分析中确定极端运行方式?均亟需研究未来高比例可再生能源等新因素接入下电力系统运行方式的变化机理。

电力系统运行方式的变化具体表现为线路潮流、机组出力、节点电压、相角等相关电气量运行规律的变化。实际电力系统中包含大量的元件(线路、机组、母线),多样的电气量(有功、无功、电压、电流、相角),不同的时间尺度(年、月、日)和时间粒度(小时级、分钟级、秒级),因此运行方式数据是高维且非线性相关的。例如江苏省 220kV 以上线路超过 2500 条,如果仅考虑小时级时间粒度,则全年存在 365 个 60000 维日潮流向量。这些向量内的不同维度之间存在时空及潮流方程耦合。因此,电力系统运行方式分析是典型的高维非线性数据分析问题。

学术界已经对数据驱动技术在电力系统中的应用进行了初步探索<sup>[10]</sup>。这些研究主要集中在 3 个方面:预测,模式识别和决策分析。预测主要包括负荷预测<sup>[11-12]</sup>和可再生能源预测<sup>[13-15]</sup>,基本思路是通过数据驱动的方法建立负荷或可再生能源出力与影响因素之间的相关模型而非因果模型进行预测,为负荷预测提供了新的思路。文献[12]先将负荷聚类分组,再通过集成学习的方法预测总负荷。文献[11]则将 LSTM 模型应用于个体概率性负荷预测。文献[15]通过深度学习的方法选取主要特征以提高风电短期出力预测精度。在模式识别方面,数据驱动的方法主要用于电力系统状态估计<sup>[16]</sup>,拓扑识别<sup>[17-18]</sup>,和事故辨识<sup>[19-21]</sup>等。文献[16]通过数据驱动方法对历史数据进行学习以改善智能电网在线状态估计的效果。在输入输出量测均存在误差的情况下,文献[18]提出了一种可以同时估计线路参数和拓扑的数据驱动方法。文献[19]基于 PMU 数据,利用数据驱动方法建立低维数据空间的线下字典,并应用于线上事故实时检测。文献[21]则利用数据驱动的聚类技术对 SCADA 系统的潜在攻击进行检测。在决策分析方面,数据驱动的方法主要应用于电力系统优化运行<sup>[22-24]</sup>,规划决策<sup>[25-29]</sup>,以及需求侧响应<sup>[30]</sup>等。文献[22]利用斜决策树学习电力系统的安全规则,并将安全规则转换为线性约束嵌入最优潮流模型以指导电力系统安全优化调度,实验结果表明该方法保证了日前调度结果 95% 以上为安全状态。文献[25]结合聚类、特征提取和 copula

技术提出了一种新的场景生成方法,实验结果证明了提出的方法应用于规划可以达到近似最优的投资决策。文献[27]提出了一种分层聚类技术选取典型场景,并且可以考虑场景间的时空相关性。文献[28]利用数据驱动的方法对用户行为模式进行表征,进而提出了考虑社会和技术复杂性的分布式能源规划决策模型。文献[29]提出了一种可以考虑可再生能源出力波动性的运行状态聚类方法。文献[30]提出了一种通过数据驱动模型为用户进行需求侧响应决策的方法,该方法比基于模型的方法效益提高了 17%。文献[31]利用长短期记忆网络等数据驱动方法实现了从负荷到调度决策结果的端到端映射以快速求解机组组合问题。文献[32]综述了利用人工智能技术进行电力系统暂态响应分析和控制决策的方法及挑战。

总的来说,已经有许多数据驱动的方法应用于运行规划过程中电力系统典型运行场景选取,但是缺少高比例可再生能源等新因素接入下电力系统运行方式及其模式变化规律的研究。在实际电力系统规划中,仍然沿用“冬大冬小,夏大夏小,丰水枯水”等通过经验选取电力系统运行方式数量及模式的方法。为此,本文基于电力系统精细化运行模拟仿真数据,提出了数据驱动的电力系统运行方式分析方法及框架;该方法首先对海量高维运行模拟数据进行预处理以压缩其中的冗余维度,提高后续算法的运行效率;接着利用聚类算法和紧密度指标饱和点分别在高维空间发现典型运行模式及其数量;然后通过降维可视化算法提取运行方式主要特征并在低维空间直观展示运行方式的高维空间分布及时序特点;接着,从定量的角度,本文提出了描述电力系统运行方式分散性、季节一致性、时序多变性等特点的数学指标。最后,以甘肃电网实际算例实证分析了高比例可再生能源对电力系统运行方式影响,验证了本文提出的框架和数据驱动方法的有效性。

本文的创新性主要体现在以下几个方面:

1) 精细化运行模拟生成海量运行方式数据:数据驱动的方法需要足够的支撑,但是电力系统实验成本高昂且部分研究场景在当前实际系统中并不存在,例如超高比例可再生能源电力系统,本文的方法基于精细化运行模拟数据产生海量运行方式数据,以解决实际电力系统中历史运行方式数据不足的问题;

2) 数据驱动的新视角:通过紧密度指标饱和

点和聚类等数据驱动算法分析电力系统运行方式数据，从而辨识高比例可再生能源、储能等新因素接入下电力系统典型运行方式及其数目，避免了基于经验选取典型运行方式的方法在筛选海量运行场景方面的不足；

3) 从定性和定量两个方面展示电力系统运行方式的特点：通过高维聚类和降维可视化，在低维展示了电力系统运行方式的高维空间分布及时序变化，能够帮助我们直观分析和理解新因素对电力系统运行方式的影响，同时通过对运行方式直观特点的定量描述，提出了运行方式分散性、季节一致性、时序多变性等严格指标，有助于我们将所得结论应用于电力系统运行规划；

4) 适用于实际电力系统：电力系统数据具有高维复杂非线性相关的特点，特别是对于大型实际电力系统，本文通过甘肃省实际算例验证了所提方法的有效性并对比分析了不同可再生能源接入下甘肃电网运行方式的特点。实验结果表明电力系统运行方式的分散性及典型运行方式的数量会随着可再生能源渗透率的上升而迅速增加。此外，电力系统运行方式与季节的耦合性逐渐减弱。

本文第 1 节将介绍在新的关键因素影响下，利用数据驱动方法进行电力系统运行方式分析的框架。第 2 节将具体展开进行运行方式分析的数据驱动算法。第 3 节提出了定量描述运行方式分散性、季节一致性和时序多变性等特点的数学指标。第 4 节实证分析了高比例可再生能源接入下甘肃电网运行方式及其典型模式的变化。最后，总结全文和展望进一步的研究。需要注意的是，本文提到的可再生能源均指风电和光伏等间歇性可再生能源。

### 1 数据驱动的电力系统运行方式分析框架

电力系统运行方式数据具有高维且非线性相关的特点，利用数据驱动的方法分析电力系统的运行方式，需要解决以下 3 个问题：

- 1) 如何获取新因素作用下的海量运行方式数据？
- 2) 如何高效地对高维非线性相关的运行方式数据进行分析？
- 3) 如何定量地对电力系统运行方式的特性进行描述？

为此，图 1 给出了数据驱动的电力系统运行方式分析框架：

1) 确定边界条件：根据研究因素(可再生能源、储能等)确定电力系统的边界条件，例如，研究高比

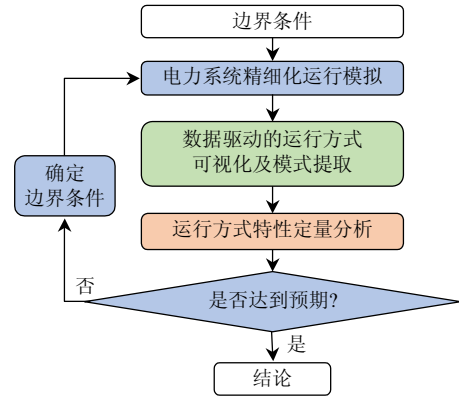


图 1 数据驱动的电力系统运行方式分析框架

Fig. 1 Data-driven framework for power system operation mode analysis

例可再生能源时需要确定可再生能源机组容量、出力概率分布、时空相关性、季日特性、线路和负荷参数等；

2) 电力系统精细化运行模拟：数据驱动的方法需要足够的支撑，但是电力系统实验成本高昂且部分研究场景在当前实际系统中并不存在，例如超高比例可再生能源电力系统，因此，本框架基于电力系统精细化运行模拟产生海量运行方式数据，以解决实际电力系统中历史运行方式数据不足的问题。图 2 给出了本文的电力系统运行模拟方法，首先通过随机微分方程重构考虑时空相关性、季日特性的可再生能源时序出力，同时进行水电中长期运行模拟和火电全年检修模拟；通过可再生能源、水电、检修模拟得到的日边界条件，以最小化运行成本为目标，逐日模拟考虑多种电源类型和网络约束的电力系统优化调度，得到电力系统机组出力，线路潮流、负荷等运行方式数据。电力系统运行模拟方法较为成熟，目前已有一些商业软件，因此具体方法不再赘述<sup>[33]</sup>。为了全面地反映电力系统“源-网-荷”的总体特征，本文基于运行模拟数据定义电力系统运行方式为

$$p = (g_{1 \times (|g| \times T_s)}, r_{1 \times (|r| \times T_s)}, f_{1 \times (|f| \times T_s)}, d_{1 \times (|d| \times T_s)})^T \quad (1)$$

式中传统机组调度出力  $g$  和可再生能源机组出力  $r$

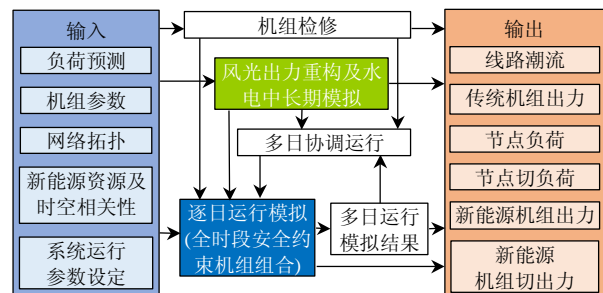


图 2 电力系统运行模拟框架

Fig. 2 Power system operation simulation framework

代表源侧运行方式数据, 线路潮流  $f$  代表网侧运行方式数据, 负荷  $d$  代表负荷侧的特征。  $T_s$  代表每天的时段,  $|g|$ 、 $|r|$ 、 $|f|$ 、 $|d|$  分别代表传统机组、可再生能源机组、线路和负荷节点的数量。需要说明的是, 面对历史运行方式后评估时, 电力系统运行方式数据也可以通过电力系统实际运行数据的采集获取。

3) 数据驱动的运行模式提取及可视化: 电力系统运行方式数据呈现海量高维且非线性相关的特点。因此, 本框架基于电力系统精细化运行模拟的结果, 首先对高维运行方式数据进行预处理, 以排除不合理的运行状态和冗余维度以提高后续算法的运行效率; 接着通过紧密度指标和聚类算法分析电力系统运行模拟数据, 辨识高比例可再生能源、储能等新因素接入下电力系统典型运行方式及其数量, 避免了基于经验选取典型运行方式的方法在筛选海量运行场景方面的不足; 最后利用 t-SNE<sup>[34]</sup>或 LargeVis<sup>[35]</sup>等降维算法解耦高维向量的相关性、提取主要特征并进行可视化。通过运行方式散点图和时序图在低维空间展示电力系统运行方式及相关特性的变化, 能够帮助我们直观分析和理解新因素对电力系统运行方式的影响。具体数据驱动的方法详见第2节。

4) 运行方式特性定量分析: 基于运行方式可视化得到的直观认知, 对运行方式的特性进行数学指标建模和定量描述, 有助于我们将所得结论应用于电力系统运行规划。本文提出运行方式分散性、季节一致性、时序多变性等指标定量描述电力系统运行方式的变化, 详见第3节。

5) 反馈: 如果达到了预期实验目的则结束分析、得出结论, 否则根据研究对象重新确定边界条件, 返回步骤2。

## 2 数据驱动方法

### 2.1 电力系统运行方式数据的特点及数据驱动算法选取原则

电力系统运行方式数据具有以下特点:

1) 数据规模大, 维度高: 电力系统运行方式的变化具体表现为线路潮流、机组出力、节点电压相角等相关电气量运行规律的变化。实际电力系统中包含大量的元件, 多样的电气量, 不同的时间尺度和时间粒度。因此其数据维度高, 数据规模大。例如江苏省 220kV 以上线路超过 2500 条, 如果仅考虑小时级时间粒度, 则全年存在 365 个 60000 维日潮流向量。此外, 随着高比例可再生能源、储能

等新因素的接入下, 电力系统运行方式呈现进一步多样化的趋势, 因此, 需要仿真更多的运行方式样本分析其规律, 这种新的特点进一步增大了所需电力系统运行方式数据的规模。如果使用电力系统历史数据, 则由于实际电力系统中广泛存在的量测设备, 仍然面临数据规模大、维度高的问题。因此, 一方面有必要对运行方式数据进行预处理以提高后续算法的效率, 另一方面高维度运行方式数据难以直观理解, 有必要通过聚类算法确定其在高维空间的聚类结构, 并通过降维可视化的方法在低维空间直观展示其高维空间分布及时序特点, 从而加深我们在新因素接入下对电力系统运行方式特点的理解。

2) 不同维度之间非线性相关: 电力系统运行方式数据的不同维度之间存在物理规律的时空耦合, 例如基尔霍夫电流定律、电压定律、潮流方程的耦合等, 因此不同维度之间存在时间和空间的复杂非线性相关性, 同时存在一定的冗余性, 例如机组出力、负荷及线路潮流数据只要知道前两者就可以通过潮流方程计算出线路潮流。因此我们有必要通过预处理算法对冗余维度进行压缩。同时需要降维可视化算法解耦其在高维空间的相关性并投影到二维平面进行直观展示。

如综上所述, 本文采用了预处理、聚类、降维可视化的思路对海量运行方式数据进行分析(如图3所示)。预处理、聚类、可视化降维等数据驱动方法应当通过以下原则选取:

预处理算法应当能快速处理大规模高维运行方式数据, 能对运行方式的冗余维度进行压缩并表示其压缩程度, 对于实际数据需要能辨识异常运行方式, 以提升后续算法的运行效率; 聚类算法应当能快速处理大规模样本, 鲁棒性高, 同时应当通过适当的指标合理确定电力系统运行方式数量; 降维可视化算法应当能够解耦高维非线性相关运行方式数据, 同时能在低维平面反映数据在高维空间的分布, 例如在高维空间聚在一类的运行方式在低维

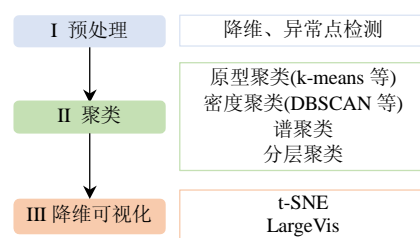


图3 数据驱动方法流程

Fig. 3 Flow chart of data-driven method

空间应当相互邻近。下面将具体介绍各环节的典型算法。

## 2.2 预处理

由于电力系统实际和仿真数据往往存在坏数据或维度冗余的问题，通常需要对实际和仿真数据进行预处理。对于仿真数据，常用的预处理方法是降维压缩。降维压缩可以过滤掉电力系统实际和仿真数据中冗余的维度，提高后续聚类及可视化算法的效率，从而将数据驱动方法拓展到实际大型电力系统。常用的降维预处理算法有主成分分析(PCA)，多维变量缩放等<sup>[36]</sup>。本文使用效率较高且易于控制压缩程度的 PCA 算法对高维运行方式数据进行降维预处理。

主成分分析的基本原理是在保持样本点中方差最大的特征的前提下，对数据进行线性变换从而实现降维。设  $M$  个  $N$  维日运行方式向量组成矩阵  $P_0=(p_1, p_2, \dots, p_M)$ 。则运行方式数据预处理的步骤为：

- 1) 对所有运行方式向量进行中心化：

$$P = (p_1 - \bar{p} \quad p_2 - \bar{p} \quad \dots \quad p_M - \bar{p}) \quad (2)$$

其中  $\bar{p}$  为运行方式向量的平均值。

- 2) 计算协方差矩阵  $Cov = PP^T / M$ ，特征值分解的得特征值  $\lambda_1, \lambda_2, \dots, \lambda_N$  和对应特征向量  $h_1, h_2, \dots, h_N$ ，其中  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ 。

- 3) 取前  $K$  个特征向量组成矩阵  $H=(h_1, h_2, \dots, h_K)$ 。其中，压缩后运行方式向量维数  $K$  可以通过压缩系数  $\theta_0$  确定， $\theta_0=0$  代表不压缩：

$$K = \arg \min_K [(\sum_{i=1}^K \lambda_i / \sum_{i=1}^N \lambda_i) \geq (1 - \theta_0)] \quad (3)$$

- 4) 线性变换得到预处理后电力系统运行方式数据  $P'=(p'_1 \quad p'_2 \quad \dots \quad p'_M)$ ：

$$P' = H^T P \quad (4)$$

## 2.3 聚类及典型运行方式提取

在新的关键因素影响下，电力系统的运行方式呈海量多变的趋势，因此需要聚类发现典型运行模式。聚类的主要方法有原型聚类(k-Means)、密度聚类(DBSCAN)、分层聚类等。本文采用原型聚类中 k-Means++ 算法，其收敛速度快，鲁棒性较好，适用于大规模实际数据<sup>[37]</sup>。但在高比例可再生能源等场景下，难以经验地选取合适的聚类数目，本文提出了利用电力系统运行方式紧密度指标进行聚类数目选取的方法。

电力系统运行方式紧密度定义为每一类中各运行方式点到聚类中心的平均距离，具体如下：

$$CP = \frac{1}{L} \sum_{i=1}^L \left( \frac{1}{|\Omega_i|} \sum_{p_j \in \Omega_i} \|p'_j - \bar{p}_i\| \right) \quad (5)$$

式中： $L$  为聚类数目； $\Omega_i$  为第  $i$  类运行方式的集合； $\bar{p}_i$  代表第  $i$  类运行方式的聚类中心。紧密度可以用于描述在一定分类方式下运行方式的分散性。一般来说紧密度随着聚类数目的上升而下降，当单位聚类数目的增加使紧密度变化小于  $\delta$  时，紧密度达到饱和，饱和点对应的  $L$  可以认为是系统最优的聚类数目，如图 4 所示。

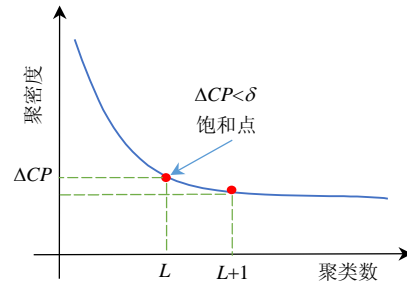


图 4 选取最优聚类数原理图

Fig. 4 Illustration of determining optimal clustering number

选取最优聚类数后，利用 k-means++ 算法对预处理后电力系统运行方式数据  $P'$  进行模式提取，其核心思想是使初始运行方式聚类中心之间的距离尽可能远，其具体步骤如下：

- 1) 随机选取一个运行方式向量作为第一个聚类中心  $\bar{p}_1$ ；

- 2) 对于剩余的  $L-1$  个聚类中心，重复以下步骤直至所有的聚类中心全部产生：

①对于每一个运行方式向量  $p'_i$ ，计算它到所有已选出聚类中心  $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_j$  的距离的最小值  $d_i = \min \|p'_i - \bar{p}_k\| \leq k \leq j$ ；

②将  $d_1, d_2, \dots, d_n$  归一化，通过随机模拟以概率  $S$  选取对应的运行方式向量  $p'_i$  作为新的聚类中心。其中， $S$  的表达式为

$$S = d_i / (\sum_{k=1}^n d_k) \quad (6)$$

因此， $d$  值较大的运行方式向量有更高的概率被选为新的聚类中心。

- 3) 选取所有  $L$  个聚类中心后通过标准 k-means 算法对运行方式向量进行聚类，得到的聚类中心即为典型运行模式。

本文的算例分析将以典型运行方式辨识为主。需要说明的是相较于典型运行方式以代表性为主体的选取标准，极端运行方式选取的标准更加多元化，例如关键线路或断面潮流越线， $N-k$  不安全等。

通过本文提出的框架,可以将运行方式的高维空间分布及聚类信息在低维空间展示,加强对电力系统运行方式特点的理解,从而选取合适的极端运行方式。基于本文提出的数据驱动框架,建议通过以下两种方式选取极端运行方式:

1) 数据驱动算法直接选取极端运行方式:部分极端运行方式具有离群或位于运行方式空间分布图边缘的特点,因此,可以基于运行模拟仿真的海量运行方式数据,进行离散点辨识选取极端运行方式;

2) 数据驱动算法协助选取极端运行方式:由于高比例可再生能源等因素作用下,电力系统运行方式海量多变,因此可以通过聚类的方式缩小搜索极端运行方式的空间,然后利用传统原则在聚类结果中选取极端运行方式。具体方法为,首先适当增大典型运行方式数量并进行日典型运行方式聚类,接着在日典型运行方式中通过传统原则筛选符合以下条件的小时级极端运行方式:①关键线路或断面潮流可能越界的运行方式,例如关键线路或断面潮流达到90%极限;②最大弃风、弃光、弃水的运行方式;③切负荷的运行方式;④备用率最小或可靠性指标最小等运行方式等。

## 2.4 降维可视化方法

电力系统运行方式数据高维且非线性相关,难以直观理解电力系统运行方式的变化模式。因此,本文通过降维可视化算法提取电力系统运行方式的主要特征,将高维数据映射到2维空间实现电力系统高维非线性运行方式可视化。典型的高维数据可视化算法有t-SNE和LargeVis,其中t-SNE已经成功应用在自然语言处理、图像处理可视化等多个领域<sup>[34,38-39]</sup>。t-SNE的基本思路是将高维运行方式数据和一个低维流形建立对应关系,并通过计算得到相应运行方式在低维流形中的具体位置。设低维运行方式数据为 $\mathbf{Y}=(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ 。

在t-SNE中,高维运行方式分布用高斯分布表达。则任意两个高维运行方式向量 $\mathbf{p}_i$ 和 $\mathbf{p}_j$ 之间的条件概率表达式 $s_{ji}$ 为

$$s_{ji} = \frac{\exp(-\frac{\|\mathbf{p}'_i - \mathbf{p}'_j\|^2}{2\sigma_i^2})}{\sum_{k \neq i} \exp(-\frac{\|\mathbf{p}'_i - \mathbf{p}'_k\|^2}{2\sigma_i^2})} \quad (7)$$

式中 $\sigma_i$ 为二维搜索得到的高斯分布项的方差项。根据公式(7),任意两个高维运行方式向量 $\mathbf{p}'_i$ 和 $\mathbf{p}'_j$ 之

间的分布 $s_{ij}$ 表示为

$$s_{ij} = \frac{s_{ij} + s_{ji}}{2M} \quad (8)$$

低维运行方式的分布用 $t$ 分布表达,则任意两个低维运行方式向量 $\mathbf{y}_i$ 和 $\mathbf{y}_j$ 之间的分布 $q_{ij}$ 为

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{m \neq n} (1 + \|\mathbf{y}_m - \mathbf{y}_n\|^2)^{-1}} \quad (9)$$

t-SNE通过最小化目标函数KL散度 $D_{KL}$ 使高维和低维运行方式分布尽可能接近以达到降维的目的:

$$D_{KL} = \sum_{i \neq j} s_{ij} \log \frac{s_{ij}}{q_{ij}} \quad (10)$$

## 3 电力系统运行方式特征指标

通过数据驱动的方法对电力系统运行方式进行聚类、降维、可视化分析可以得到对电力系统运行方式空间分布的直观理解。为了定量描述电力系统的运行方式的特征以进一步将指导电力系统运行规划,本文提出3个指标分别描述电力系统运行方式的分散性、季节的一致性以及时序多变性。

### 3.1 分散性

方差可以描述样本每一维的分散程度,本文定义运行方式向量每一维方差的平均值为高维方差指标HV,用于描述电力系统运行方式整体的分散性,表达式为

$$HV = \frac{1}{N} \text{Trace}(\mathbf{P}\mathbf{P}^T) \quad (11)$$

从式(11)可以看出,高维方差越小,运行方式整体分散性越小。

### 3.2 季节一致性

为了描述运行方式与季节的一致性,本文定义季节一致性指标SC为每个季节主导运行模式的样本数占总体样本数的比值:

$$SC = \sum_{i=1}^4 m_i / M \quad (12)$$

式中 $m_i$ 代表第 $i$ 个季节中主导运行模式的样本数。从定义及公式可以看出,季节一致性指标越大,电力系统运行方式与季节的一致性越高。

### 3.3 时序多变性

为了描述聚类后日运行方式变化的频率,本文定义日运行方式发生变化的样本数与总体样本数的比值为时序多变性CF,表达式为

$$\begin{cases} CF = \sum_{i=1}^M I(p_i) / M \\ I(p_i) = \begin{cases} 1, & p_i \in \Omega \text{ 且 } p_{i+1} \notin \Omega \\ 0, & p_i \in \Omega \text{ 且 } p_{i+1} \in \Omega \end{cases} \end{cases} \quad (13)$$

式中 $\Omega$ 为样本  $p_i$  所属的类。从式(13)可以看出，时序多变性指标越大，系统的日运行方式发生变化的频率越高。

## 4 算例分析

### 4.1 甘肃电网算例数据

为研究高比例可再生能源渗透率对电力系统运行方式的影响，本文以甘肃省电网实际数据进行实证分析。截止 2020 年上半年，甘肃省可再生能源最大出力超过 11000MW，占甘肃电网用电负荷 87.3%，占当时总发电出力的 57.2%<sup>[40]</sup>。随着可再生能源渗透率的提高，甘肃电网运行方式面临巨大挑战。本文以甘肃电网实际数据和网架结构为基准，设计了低比例(0%)，中比例(20%)，高比例(30%)等 3 种可再生能源渗透率情景，分别代表过去、当前、及未来电网的形态，如表 1 所示。其中，低比

表 1 甘肃电网算例数据

Table 1 Case study data of Gansu provincial power system

情形	火电装机/MW	水电装机/MW	风电装机/MW	光伏装机/MW	可再生能源电量渗透率/%
低比例	18260	9258	19	7	0
中比例	18260	9258	12772	6502	20
高比例	18260	9258	14901	11002	30

例算例为对比算例；为了不同方案之间的可比性，网架结构，传统机组装机等基本参数均按 2018 年考虑；可再生能源电量渗透率由全年时序运行模拟后评估得到。

### 4.2 高比例可再生能源渗透率对甘肃电网运行方式的影响

基于上述边界条件并利用甘肃可再生能源电站历史数据拟合风电、光伏出力概率分布及时空相关性，本文首先进行可再生能源时序出力重构及甘肃电网全年精细化运行模拟，并利用本文提出的数据驱动方法对高维日运行方式数据进行预处理压缩、聚类，最终降维可视化得到运行方式的二维状态空间分布及时序变化(如图 5 所示)。其中，每一个点代表一个日运行方式，不同的颜色代表聚类得到的运行模式，时序变化图中五角星代表每个运行模式的聚类中心即典型运行模式，预处理压缩系数为 0.01%。

从运行方式状态空间分布图中可以看出，典型运行模式的数量随着可再生能源比例的上升而显著增加，从低比例时的 3 个典型日运行方式，上升到中比例时 5 个，最终达到高比例时 8 个。此外，电力系统运行方式的分散性逐渐增强，同一个模式下运行方式之间的差异性增大，这意味着典型运行方式的代表性变差，在运行规划中需要考虑更多的典型及极端运行方式来代表海量场景。从运行方式时序图可以看出，随着可再生能源比例的增加，日

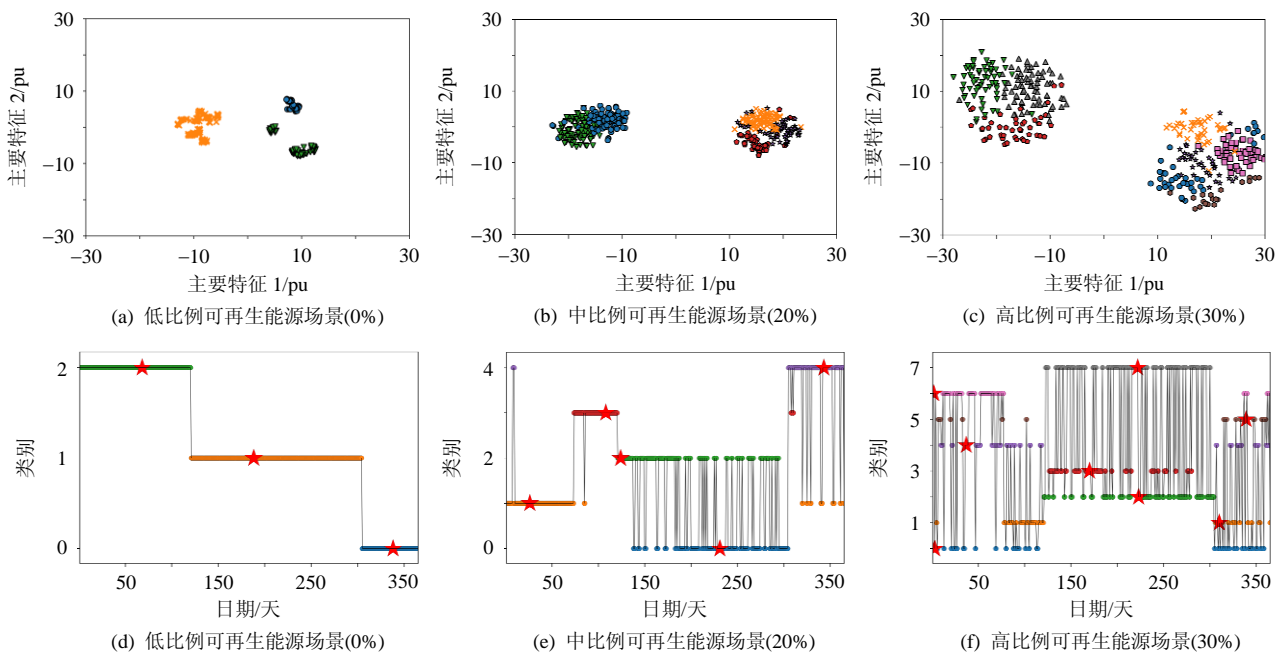


图 5 三种可再生能源场景下甘肃省电力系统运行方式空间分布及时序图

Fig. 5 Operation modes visualization of Gansu provincial power system under three renewable energy penetration scenarios

运行方式的波动性急速上升。当可再生能源比例较低时,甘肃电网的运行方式主要受季节性负荷、水电出力影响,季节内日运行方式相对稳定,因此,在运行规划中基于经验“冬大冬小,夏大夏小,丰水枯水”选取典型运行方式的传统方法仍然适用。例如,甘肃电网在低比例情形下,全年仅发生两次运行方式变化。当可再生能源比例上升时,电力系统的运行方式逐渐由风电、光伏等主导,其间歇性、强不确定性的特点使得电力系统日运行方式急剧变化。例如,甘肃电网在中比例情形下仅能保持春季日运行方式的稳定,夏秋季容易受光伏出力波动和水电调节共同作用的影响,冬季易受风电出力波动影响,当可再生能源比例进一步增加时,甘肃电网日运行方式的季节模式将完全消失,相邻日期之间的运行方式剧烈波动,因此,在高比例可再生能源渗透率下,系统需要预留更多的灵活性资源,例如储能、可调度机组来应对运行方式的变化,以保证电力系统安全稳定运行。

图6和图7进一步从定量的角度分别给出了可再生能源渗透率对甘肃电网运行方式分散性、时序多变性以及季节一致性的影响。从高维方差指标可以看出,甘肃电网运行方式的分散性先随着可再生能源的渗透率缓慢增加,渗透率从0%增加到20%时,方差仅增加33%。然而,当渗透率从20%增加到30%,方差则上升了一倍以上,运行方式分散性明显增加。从时序多变性指标可以看出,低比例时全年日运行方式仅变化两次,但高比例时则全年55%的天数会发生运行方式模式变化(200次),其总体的变化趋势与分散性一致,当渗透率超过阈值20%,运行方式变化频率急剧增加。图7则给出了可再生能源渗透率对甘肃电网运行方式季节一致

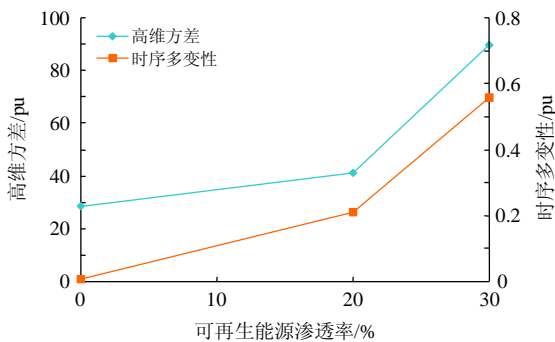


图6 可再生能源渗透率对甘肃省电力系统运行方式分散性和时序多变性的影响

Fig. 6 Impact of renewable energy penetration on Gansu provincial power system operation modes dispersion and variation

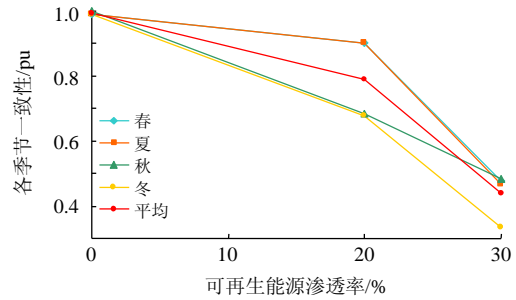
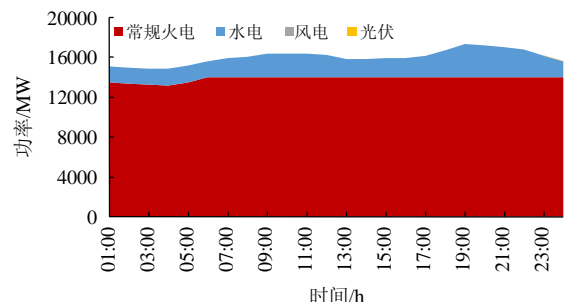


图7 可再生能源渗透率对甘肃省电力系统运行方式与季节一致性的影响

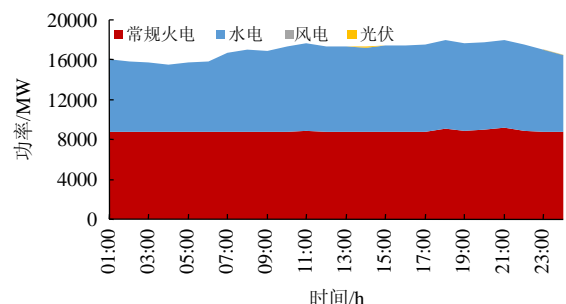
Fig. 7 Impact of renewable energy penetration on Gansu provincial power system operation modes seasonal consistency

性的影响。从图中可以看出,甘肃电网秋冬季的运行方式一致性明显低于春夏两季,在中高比例情形下,需要尤其注意冬季的运行方式变化情况。

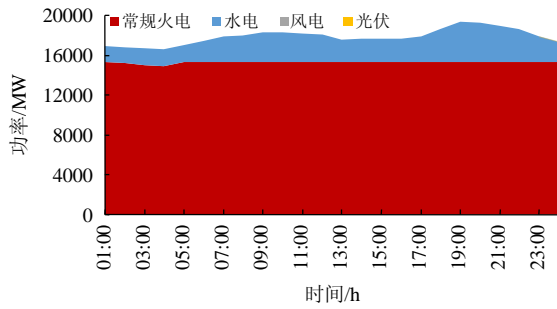
图8和图9分别给出了由聚类中心得到的低比例情形下甘肃电网的3种运行模式以及高比例情形下的8种运行模式。在低比例情形下,甘肃电网的3种运行模式分别为:小负荷模式,大负荷模式以及水电大发模式。因此,电力系统的运行方式主要由负荷和水电主导,呈现明显的季节性。当可再生能源占比上升为30%时,除了季节性水电和负荷,甘肃电网的运行方式更多的被风光主导,包含风光大发或小发等多种组合模式。由于风电和光伏在日内出力的波动性,水电需要调节出力以提高可再生能源消纳能力,因此水电也呈现出更多的日运行模式,例如1月3日水电在后半天给风电腾出力空间。此外,高比例可再生能源渗透率下甘肃电网在相邻



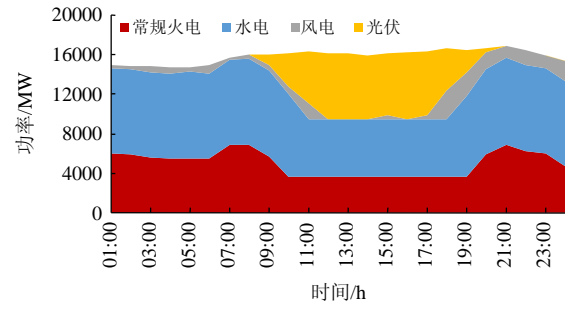
(a) 3月9日小负荷水电小发



(b) 7月7日水电大发模式



(c) 12月4日大负荷水电小发

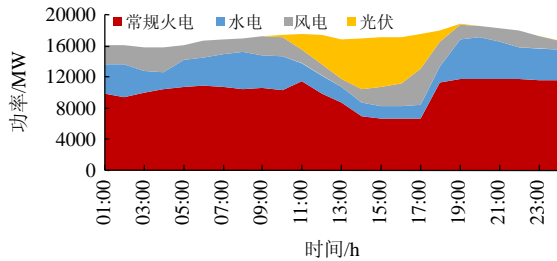


(e) 8月10日小负荷, 光水大发

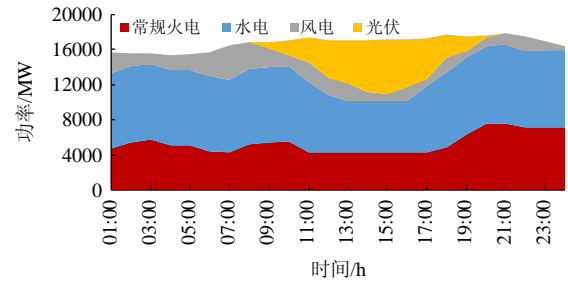
图 8 低比例可再生能源渗透率下  
甘肃省电力系统运行模式

Fig. 8 Gansu provincial power system operation patterns  
under low renewable energy penetration

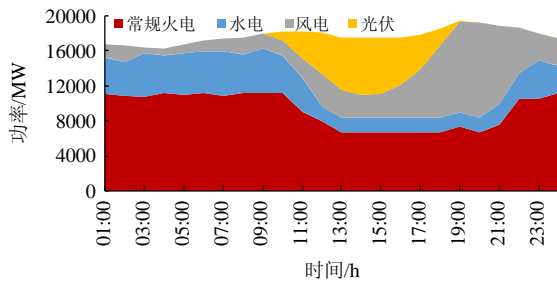
两天可能呈现完全不同的运行模式, 例如 1 月 2 日和 1 月 3 日, 8 月 10 日和 8 月 11 日。因此, 高比例可再生能源电力系统的运行方式呈多样化, 多变



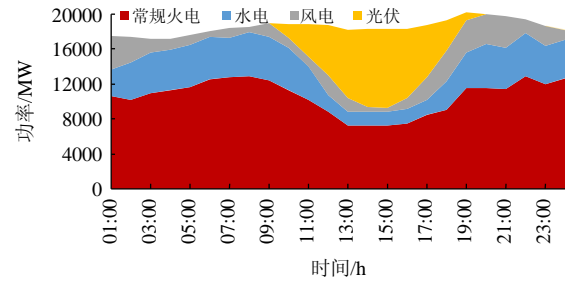
(a) 1月2日风光小发, 水电常发



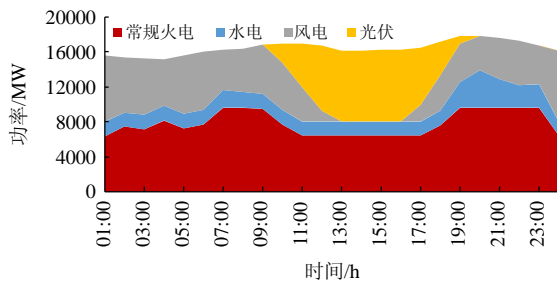
(f) 8月11日水电大发, 风光小发



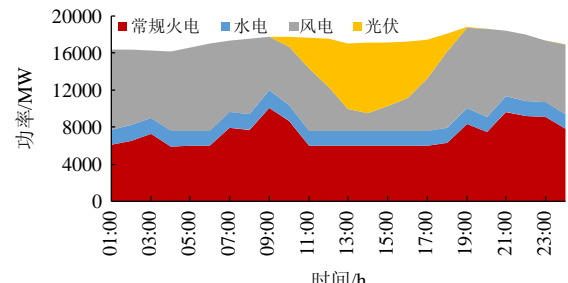
(b) 1月3日风电夜间大发



(g) 11月6日大负荷模式



(c) 2月6日风光大发, 水电小发



(h) 12月5日风电大发, 光水小发

图 9 高比例可再生能源渗透率下  
甘肃省电力系统运行模式

Fig. 9 Gansu provincial power system operation patterns  
under high renewable energy penetration

性的趋势。

从以上分析可以看出, 基于运行模拟数据和数据驱动方法, 可以同时电力运行方式进行定性理解和定量分析, 有效发掘其中的运行模式和规律。对于甘肃电网, 随着可再生能源渗透率上升, 系统的运行方式逐渐由风光主导, 运行方式分散性增强, 时序波动性增加, 季节性负荷和水电对运行方式的主导作用逐渐下降, 因此, 需要在规划中选取足够数量的典型运行方式进行经济性分析和安全校核, 在运行中关注特定季节的运行方式变化规

律,预留足够的灵活性资源以应对日运行方式的剧烈变动,在稳定和保护分析中需要分析更多的极端运行方式并制定应对措施,以保证电力系统经济、安全、稳定运行。

## 5 结论

高比例可再生能源等新因素接入使得电力系统运行方式呈现多样化的趋势。本文基于电力系统精细化运行模拟仿真数据,提出了数据驱动的电力系统运行方式分析方法。该方法利用预处理、聚类、降维可视化算法直观分析和定性理解电力系统运行方式及其变化规律,接着设计定量指标描述电力系统运行方式及其变化的特点,以指导实际电力系统运行规划。通过甘肃省电网的实际算例,实证分析了高比例可再生能源对电力系统运行方式的影响,本文主要结论如下:

1) 可再生能源渗透率较低时,电力系统中不同模式的运行方式差别明显、同模式的运行方式紧密聚集,随着可再生能源渗透率的上升,电力系统运行方式的方差逐渐增大,不同模式的运行方式出现交叠、同模式运行方式的分散性逐渐增强,这说明电力系统的运行方式呈现多样化的趋势。

2) 在可再生能源渗透率较低时,电力系统运行方式基本与季节的变化趋势重合,传统规划中“冬大冬小,夏大夏小,丰水枯水”选取典型运行方式的原则具有较好地效果,但随着可再生能源渗透率的上升,电力系统运行方式与季节的相关性减弱,运行模式数量上升,需要更多的典型运行方式来还原真实的情况。

3) 随着可再生能源渗透率的上升,电力系统运行方式的变化频率显著增加,相邻两天的运行方式可能呈现完全不同的模式,这说明电力系统运行调度中需要更多的灵活性资源和可调度手段,以应对运行方式的频繁变化带来的功率平衡问题。

4) 无可再生能源或可再生能源渗透率较低时电力系统的运行方式由负荷和水电主导,当可再生能源渗透率超过某一个阈值,系统的运行方式及其模式逐渐由可再生能源主导,分散性增强,变化频率上升,季节性消失,深刻改变了电力系统的运行稳定机理。在稳定和保护分析中需要辨识更多的极端运行方式并制定应对措施,以保证电力系统经济、安全、稳定运行。

## 6 展望

本文重点对高比例可再生能源接入下电力系统运行方式及其变化规律进行了定性和定量的分析。但电网中各类新的要素不断增加,例如大量储能以及电动汽车逐渐接入电网。电热气等多种能源系统的耦合增加,电力系统的运行方式和形态可能发生新的变化。因此,未来的研究可以从以下两个方面展开:

1) 利用本文提出的数据驱动方法分析高比例集中式或分散式储能的接入电网以及多能源系统耦合等因素对电力系统运行方式的影响。

2) 将本文的结论应用于实际电力系统运行规划中选取典型和极端运行方式。

## 参考文献

- [1] 国家发改委能源研究所. 中国 2050 高比例可再生能源发展情景暨路径研究[J]. 北京: 国家发改委能源研究所, 2015.  
Institute of energy research of national development and reform commission. China 2050 high renewable energy penetration scenario and roadmap study[J]. Beijing: Institute of energy research of national development and reform commission, 2015(in Chinese).
- [2] KROPOSKI B, JOHNSON B, ZHANG Yingchen, et al. Achieving a 100% renewable grid: operating electric power systems with extremely high levels of variable renewable energy[J]. IEEE Power and Energy Magazine, 2017, 15(2): 61-73.
- [3] 康重庆, 姚良忠. 高比例可再生能源电力系统的关键科学问题与理论研究框架[J]. 电力系统自动化, 2017, 41(9): 2-11.  
KANG Chongqing, YAO Liangzhong. Key scientific issues and theoretical research framework for power systems with high proportion of renewable energy[J]. Automation of Electric Power Systems, 2017, 41(9): 2-11(in Chinese).
- [4] 袁小明, 程时杰, 胡家兵. 电力电子化电力系统多尺度电压功角动态稳定问题[J]. 中国电机工程学报, 2016, 36(19): 5145-5154.  
YUAN Xiaoming, CHENG Shijie, HU Jiabing. Multi-time scale voltage and power angle dynamics in power electronics dominated large power systems[J]. Proceedings of the CSEE, 2016, 36(19): 5145-5154(in Chinese).
- [5] HOU Qingchun, ZHANG Ning, DU Ershun, et al. Probabilistic duck curve in high PV penetration power system: Concept, modeling, and empirical analysis in China[J]. Applied Energy, 2019, 242: 205-215.

- [6] 鲁宗相, 李海波, 乔颖. 含高比例可再生能源电力系统灵活性规划及挑战[J]. 电力系统自动化, 2016, 40(13): 147-158.  
LU Zongxiang, LI Haibo, QIAO Ying. Power system flexibility planning and challenges considering high proportion of renewable energy[J]. Automation of Electric Power Systems, 2016, 40(13): 147-158(in Chinese).
- [7] HOU Qingchun, YU Yanghao, DU Ershun, et al. Embedding scrapping criterion and degradation model in optimal operation of peak-shaving lithium-ion battery energy storage[J]. Applied Energy, 2020, 278: 115601.
- [8] HOU Qingchun, DU Ershun, ZHANG Ning, et al. Impact of high renewable penetration on the power system operation mode: a data-driven approach[J]. IEEE Transactions on Power Systems, 2020, 35(1): 731-741.
- [9] 薛禹胜, 赖业宁. 大能源思维与大数据思维的融合(一) 大数据与电力大数据[J]. 电力系统自动化, 2016, 40(1): 1-8.  
XUE Yusheng, LAI Yening. Integration of macro energy thinking and big data thinking part one big data and power big data[J]. Automation of Electric Power Systems, 2016, 40(1): 1-8(in Chinese).
- [10] 薛禹胜, 赖业宁. 大能源思维与大数据思维的融合(二) 应用及探索[J]. 电力系统自动化, 2016, 40(8): 1-13.  
XUE Yusheng, LAI Yening. Integration of macro energy thinking and big data thinking part two applications and explorations[J]. Automation of Electric Power Systems, 2016, 40(8): 1-13(in Chinese).
- [11] WANG Yi, GAN Dahua, SUN Mingyang, et al. Probabilistic individual load forecasting using pinball loss guided LSTM[J]. Applied Energy, 2019, 235: 10-20.
- [12] WANG Yi, CHEN Qixin, SUN Mingyang, et al. An ensemble forecasting method for the aggregated load with subprofiles[J]. IEEE Transactions on Smart Grid, 2018, 9(4): 3906-3908.
- [13] LI Gong, SHI Jing. On comparing three artificial neural networks for wind speed forecasting[J]. Applied Energy, 2010, 87(7): 2313-2320.
- [14] GHORBANI M A, KHATIBI R, FAZELIFARD M H, et al. Short-term wind speed predictions with machine learning techniques[J]. Meteorology and Atmospheric Physics, 2016, 128(1): 57-72.
- [15] FENG Cong, CUI Mingjian, HODGE B M, et al. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting[J]. Applied Energy, 2017, 190: 1245-1257.
- [16] WENG Yang, NEGI R, FALOUTSOS C, et al. Robust data-driven state estimation for smart grid[J]. IEEE Transactions on Smart Grid, 2017, 8(4): 1956-1967.
- [17] ZHANG Jiawei, WANG Yi, WENG Yang, et al. Topology identification and line parameter estimation for non-PMU distribution Network: a numerical method[J]. IEEE Transactions on Smart Grid, 2020, 11(5): 4440-4453.
- [18] YU Jiafan, WENG Yang, RAJAGOPAL R. PaToPa: A data-driven parameter and topology joint estimation framework in distribution grids[J]. IEEE Transactions on Power Systems, 2018, 33(4): 4335-4347.
- [19] LI Wenting, WANG Meng, CHOW J H. Real-time event identification through low-dimensional subspace characterization of high-dimensional synchrophasor data [J]. IEEE Transactions on Power Systems, 2018, 33(5): 4937-4947.
- [20] ZHOU Yuxun, ARGHANDEH R, SPANOS C J. Partial knowledge data-driven event detection for power distribution networks[J]. IEEE Transactions on Smart Grid, 2018, 9(5): 5152-5162.
- [21] ALMALAWI A, FAHAD A, TARI Z, et al. An efficient data-driven clustering technique to detect attacks in SCADA systems[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(5): 893-906.
- [22] HOU Qingchun, ZHANG Ning, KIRSCHEN D S, et al. Sparse oblique decision tree for power system security rules extraction and embedding[J]. IEEE Transactions on Power Systems, 2020, doi: 10.1109/TPWRS.2020.3019383.
- [23] CREMER J L, KONSTANTELOS I, TINDEMANS S H, et al. Data-driven power system operation: exploring the balance between cost and risk[J]. IEEE Transactions on Power Systems, 2019, 34(1): 791-801.
- [24] HALILBAŠIĆ L, THAMS F, VENZKE A, et al. Data-driven security-constrained AC-OPF for operations and markets[C]//2018 Power Systems Computation Conference. Dublin, Ireland: IEEE, 2018: 1-7.
- [25] SUN Mingyang, CREMER J, STRBAC G. A novel data-driven scenario generation framework for transmission expansion planning with high renewable energy penetration[J]. Applied Energy, 2018, 228: 546-555.
- [26] PONCELET K, HÖSCHLE H, DELARUE E, et al. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems[J]. IEEE Transactions on Power Systems, 2017, 32(3): 1936-1948.
- [27] LIU Yixian, SIOSHANSI R, CONEJO A J. Hierarchical clustering to find representative operating periods for capacity-expansion modeling[J]. IEEE Transactions on Power Systems, 2018, 33(3): 3029-3039.
- [28] JAIN R K, QIN Junjie, RAJAGOPAL R. Data-driven planning of distributed energy resources amidst socio-

- technical complexities[J]. *Nature Energy*, 2017, 2(8): 17112.
- [29] FITIWI D Z, DE Cuadra F, OLMOS L, et al. A new approach of clustering operational states for power network expansion planning problems dealing with RES(renewable energy source) generation operational variability and uncertainty[J]. *Energy*, 2015, 90: 1360-1376.
- [30] BEHL M, SMARRA F, MANGHARAM R. DR-Advisor: A data-driven demand response recommender system[J]. *Applied Energy*, 2016, 170: 30-46.
- [31] 杨楠, 叶迪, 林杰, 等. 基于数据驱动具有自我学习能力的机组组合智能决策方法研究[J]. *中国电机工程学报*, 2019, 39(10): 2934-2945.
- YANG Nan, YE Di, LIN Jie, et al. Research on data-driven intelligent security-constrained unit commitment dispatching method with self-learning ability [J]. *Proceedings of the CSEE*, 2019, 39(10): 2934-2945 (in Chinese).
- [32] 汤奕, 崔晗, 李峰, 等. 人工智能在电力系统暂态问题中的应用综述[J]. *电机工程学报*, 2019, 39(1): 2-13.
- TANG Yi, CUI Han, LI Feng, et al. Review on artificial intelligence in power system transient stability analysis [J]. *Proceedings of the CSEE*, 2019, 39(1): 2-13 (in Chinese).
- [33] ZHANG Ning, LU Xi, MCELROY M B, et al. Reducing curtailment of wind electricity in China by employing electric boilers for heat and pumped hydro for energy storage[J]. *Applied Energy*, 2016, 184: 987-994.
- [34] Van Der Maaten L, HINTON G. Visualizing data using t-SNE[J]. *Journal of machine learning research*, 2008, 9(11): 2579-2605.
- [35] TANG Jian, LIU Jingzhou, ZHANG Ming, et al. Visualizing large-scale and high-dimensional data[C]// *Proceedings of the 25th International Conference on World Wide Web*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016: 287-297.
- [36] Van Der Maaten L, POSTMA E, Van Den Herik J. Dimensionality reduction: a comparative review[J]. *Journal of Machine Learning Research*, 2009, 10(1-41): 66-71.
- [37] UTSUNOMIYA M, Hashimoto M, Kajitani H. K-means++: the advantages of careful seeding[J]. *NEC Research and Development*, 1996, 37(3): 369-381.
- [38] VAN Der Maaten L. Accelerating t-SNE using tree-based algorithms[J]. *Journal of Machine Learning Research*, 2014, 15(93): 3221-3245.
- [39] Lecun Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [40] 甘肃新能源发电首次突破 1100 万千瓦[EB/OL]. [2020-06-21]. [http://www.gansu.gov.cn/art/2020/3/19/art\\_36\\_449312.html](http://www.gansu.gov.cn/art/2020/3/19/art_36_449312.html).
- Variable renewable energy generation in Gansu reaches 11000MW[EB/OL]. [2020-06-21]. [http://www.gansu.gov.cn/art/2020/3/19/art\\_36\\_449312.html](http://www.gansu.gov.cn/art/2020/3/19/art_36_449312.html)(in Chinese).



侯庆春

在线出版日期: 2020-10-15.

收稿日期: 2020-06-28.

作者简介:

侯庆春(1993), 男, 博士研究生, 主要研究方向为高比例可再生能源, 数据驱动的电力系统优化及分析, 电力系统优化运行及规划, [hqc16@mails.tsinghua.edu.cn](mailto:hqc16@mails.tsinghua.edu.cn);

\*通信作者: 杜尔顺(1992), 男, 博士后, 主要从事新能源发电、太阳能光热、电力规划大数据、低碳电力的研究工作, [duershun@tsinghua.edu.cn](mailto:duershun@tsinghua.edu.cn);

田旭(1990), 男, 工程师, 长期从事电力系统规划设计、新能源并网发电研究工作, [13897640106@163.com](mailto:13897640106@163.com);

刘飞(1988), 男, 硕士, 高级工程师, 研究方向为电力系统规划设计、新能源并网发电等, [liufei602@163.com](mailto:liufei602@163.com);

张宁(1985), 男, 博士, 副教授, 研究方向为可再生能源、电力系统规划与运行、多能源系统等, [ningzhang@tsinghua.edu.cn](mailto:ningzhang@tsinghua.edu.cn);

康重庆(1969), 男, 教授, 博士生导师, IEEE Fellow, 研究方向为电力规划与运行、低碳电力技术、负荷预测、电力经济等, [cqkang@tsinghua.edu.cn](mailto:cqkang@tsinghua.edu.cn).

(责任编辑 乔宝榆)

# Data-driven Power System Operation Mode Analysis

HOU Qingchun<sup>1</sup>, DU Ershun<sup>1\*</sup>, TIAN Xu<sup>2</sup>, LIU Fei<sup>2</sup>, ZHANG Ning<sup>1</sup>, KANG Chongqing<sup>1</sup>

(1. Tsinghua University; 2. State Grid Economic and Technological Research Institute of Qinghai Electric Power Company)

**KEY WORDS:** data-driven; high renewable energy penetration; power system operation mode; power system simulation; dimension reduction; clustering

The power system operation mode changes substantially due to increasing renewable energy penetration and power electronic devices integration in power system. New operation modes are emerging in power system operation, planning, protection, and stability analysis, which can hardly be handled by traditional model and experience-driven method. This paper proposes a data-driven method to analyze power system operation mode and its variation as well as identify typical operation patterns and its number based on high dimensional simulated chronological power system operation data.

The framework of the proposed method is shown in Fig. 1. Based on boundary conditions, the power system operation under high renewable energy penetration is simulated to obtain sufficient power system operation data. Then, the high-dimensional and correlated power system operation data are analyzed by the proposed data-driven pipeline. In the pipeline, we first preprocess the data by using a dimension reduction algorithm like PCA to

compress the redundant features and accelerate the following algorithm. Then, we identify representative operation mode patterns and its number by using the clustering algorithm and saturation point of compactness index, respectively. Next, a visualization algorithm like t-SNE is used to decouple the power system operation data correlation and extract key features for operation modes visualization to obtain an intuitive understanding of power system operation mode patterns and its variation. Based on the qualitative analysis, several mathematical indices such as space dispersion, seasonal consistency, and time variation are proposed to quantitatively evaluate the characteristic of power system operation modes to guide power system power operation and planning under high renewable energy penetration.

A real-world case study on Gansu provincial power system in China shows the impacts of high renewable energy penetration on power system and validates the proposed data-driven method, as shown in Fig. 2. The results show that the space dispersion and pattern number of operation modes are significantly increasing as renewable energy penetration increases. The operation modes are also less correlated with season and more likely to change its pattern in adjacent days under high renewable energy penetration. Therefore, more operation patterns should be evaluated in power system planning and more flexible resources are required in power system operation to ensure power system security.

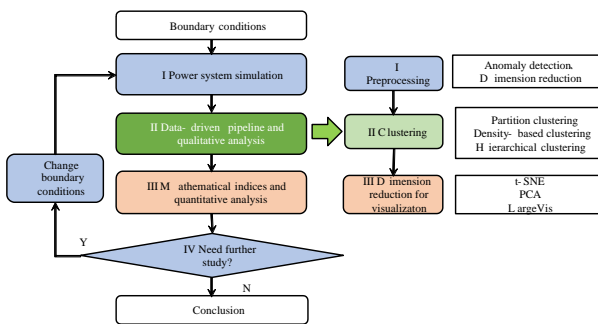


Fig. 1 Data-driven framework and pipeline

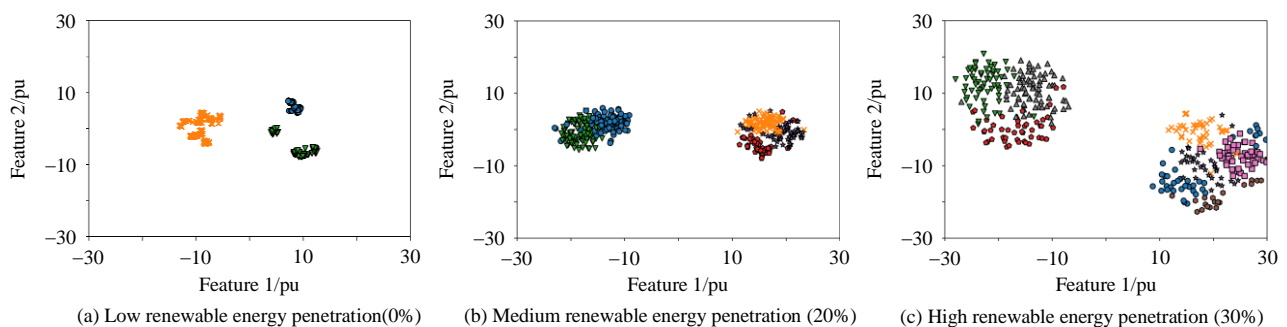


Fig. 2 Operation modes visualization of Gansu Provincial power system under three renewable energy penetration scenarios