团体标准

发 布

中国电机工程学会

20XX—XX—XX实施

20XX—XX—XX发布

电力物联网数据-机理融合建模

数据接口规范

Data interface specifications for mechanism and data dual driven modeling of EIoT

（征求意见稿）

T/CSEE XXXX—YYYY

代替 T/XXXX

ICS 19.020

CCS K85

目 次

[前 言 3](#_Toc119674377)

[1 范围 4](#_Toc119674378)

[2 规范性引用文件 4](#_Toc119674379)

[3 术语和定义 4](#_Toc119674380)

[4 总则 5](#_Toc119674381)

[5 融合建模的数据分类及要求 5](#_Toc119674382)

[5.1 数据特征 5](#_Toc119674383)

[5.2 数据分类 5](#_Toc119674384)

[5.3 数据要求 6](#_Toc119674385)

[6 数据预处理技术及要求 7](#_Toc119674386)

[6.1 数据预处理基本技术 7](#_Toc119674387)

[6.2 预处理后的数据质量要求 8](#_Toc119674388)

[附　录　A 数据-机理融合建模信息数据接口建议 10](#_Toc119674389)

前 言

本文件按照《中国电机工程学会标准管理办法（暂行）》的要求，依据GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国电机工程学会提出。

本文件由中国电机工程学会电力信息化专业委员会技术归口和解释。

本文件起草单位：清华大学、中国电力科学研究院有限公司、华北电力大学、重庆大学、国网天津市电力公司、国网江苏省电力有限公司、天津大学、国网智能电网研究院有限公司。

本文件主要起草人：、、、。

本文件首次发布。

本文件在执行过程中的意见或建议反馈至中国电机工程学会标准执行办公室（地址：北京市西城区白广路二条1 号，100761，网址：http：//www.csee.org.cn，邮箱：cseebz@csee.org.cn）。

电力物联网数据-机理融合建模数据接口规范

1. 范围

本文件规定了电力物联网数据-机理融合建模的数据接口要求，包括融合建模数据说明、数据分类、数据范围、数据预处理、预处理后的质量管理。

本文件适用于电力物联网建模相关工作，现阶段主要用于指导基于电力物联网的源网荷储协同中的数字建模工作。

1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 33474—2016 物联网 参考体系结构

GB/T 32908—2016 非结构化数据访问接口规范

GB/T 33745—2017 物联网 术语

GB/T 36478（所有部分） 物联网 信息交换和共享

GB/Z 25320.7—2015 电力系统管理及其信息交换 数据和通信安全 第7部分：网络和系统管理（NSM）的数据对象模型

YD/T 2437—2012 物联网总体框架与技术要求

YDB 100—2012 物联网需求

DL/T 1732—2017 电力物联网传感器信息模型规范

IEEE 2413—2019 IEEE Standard for an Architectural Framework for the Internet of Things (IoT)

1. 术语和定义

下列术语和定义适用于本文件。

机理驱动建模 mechanism driven modeling

根据建模对象的应用场合和模型的使用目的进行合理的假设后，依据其内在机理建立数字模型的方法。

 数据驱动建模 data driven modeling

根据建模对象的应用场合和模型的使用目的，依据可表征其状态、特性、运行环境等影响因素的历史数据和实时数据，采用统计、数据挖掘、数据解析、机器学习等数据分析手段建立数字模型的方法。

 数据-机理融合建模 mechanism and data dual driven modeling

采用机理驱动方法和数据驱动方法相结合的技术手段，通过不同机理驱动和数据驱动融合建模策略，面向建模对象建立数字模型的方法。

 外部特性模型 external characteristic model

在建模对象外部接口参数已知基础上，对若干物理参数或计算参数之间的关系进行建模。

 运行控制模型 operation control model

在建模对象部分控制逻辑和控制参数明确的情况下，对满足更多运行场景和运行目标下的控制方法进行建模。

 行为策略模型 action strategy model

在建模对象在面对不同外界环境状态时，对其具有一定随机性和不确定的行为模式和行动策略进行建模。

1. 总则

针对目前电力物联网数据-机理融合建模应用中建模数据结构复杂化、数据格式多元化、数据质量难保障等问题，可通过规范数据接口，统一数据形式，修正数据缺陷，为各类数据应用提供可靠数据基础，实现跨领域与跨业务场景数据应用的协同与共享。

融合建模数据接口应从数据需求与数据应用技术角度，明确数据特征、数据类型、数据预处理技术及质量要求，规范数据的正确性、完整性、一致性，保障数据初期获取与后期应用需求相协调。

1. 融合建模的数据分类及要求
	1. 数据特征

电力物联网融合建模数据应满足如下特征：

1. 建模数据获取频率高、类型多，既有结构化和半结构化的传感数据，也有非结构化数据。
2. 融合建模涉及的大规模、高维度数据必须保证不同数据维度之间非线性相关。
3. 根据生产流程各环节的特性，数据属性可以是实时、半实时、离线等。
4. 建模数据应具有较强的时序性。
5. 建模数据应具有闭环性，需要支撑状态感知、分析、反馈、控制等闭环场景下的动态持续调整和优化。
6. 数据驱动的方法需要大规模数据支撑，需要与机理融合建立较精细的电力系统模型，通过仿真技术生成足够的样本数据。
	1. 数据分类

根据电力物联网不同的应用需求，将建模数据分为3种类型：外部特性模型数据、运行控制模型数据、行为决策模型数据。

1. 外部特性模型数据，以能够反映建模对象运行状态的运行数据为主，包括反映外部环境的动态运行数据和反映建模对象具体类别及运行状态的静态属性数据。
2. 运行控制模型数据，包括环境变量和控制变量两部分，环境变量应合理反映建模对象的运行目标，控制变量应符合建模对象的现实约束条件。
3. 行为决策模型数据，应考虑系统参与方自身固有的行为特点以及在不同激励政策、市场行情和个体利益下的响应惯性。

其中，每类模型数据可以分别从“源-网-荷-储”架构出发作进一步限定。

1.



图1 电力物联网数据-机理融合建模分类

* 1. 数据要求
		1. 外部特性模型的数据要求

本类别用于限定电力物联网中电参量以外的表征和影响外部特性的建模数据。

1. 电源侧：对于光伏和风力发电，建模数据应包括电站所在区域的实时气象数据，例如光照强度、受光面积、切入风速、额定风速、切出风速等；对于水电，建模数据应包括流域实时水文情况与历史统计数据，例如枯水期与丰水期上游来水情况等；对于火电，建模数据应包括所在区域排污等级限制，例如二氧化硫、烟尘及氮氧化物排放量；对于沿海核电，建模数据应包括临近海域台风、海啸等极端天气预警信息；其他发电类型可忽略本类型数据。
2. 电网侧：对于输电线路，建模数据应包括沿线地区的气象环境、线路的预警信息，主要有气象、覆冰、污秽、异物、违建、周围树木、导线温度、舞动、绝缘子闪络、微风振动、通道异物、弧垂、导线温度、杆塔倾斜等参量；对于配电网，建模数据应涵盖运行环境、计量数据和用户数据，要求特别区分不同地理位置的配电物联网接入数据种类，应有助于故障快速识别和处理，应包含配电网故障点和停电地理分布，人员技能与可用物料等；对于变电站，建模数据应包括变电站环境、安防及防火系统监测数据等；对于二次设备，建模数据应包括设备所处环境的温度、湿度、通风情况、防火信息监测、空调运行工况等。
3. 负荷侧：建模数据应能反映用户用电的基本信息，包括用户用电行为、用户信息、用户地域信息、用电政策等；用电数据应包含需求侧响应数据，如用户用电类型报告，帮助用户了解和改善自身用能习惯；用户级分布式能源数据包括屋顶光伏、小型风电、生物质能、燃料电池等发电数据。
4. 储能系统：建模数据应计及系统所属地区一定范围内的实时气象信息、电池特性参数等。
	* 1. 运行控制模型的数据要求

本类别用于限定表征系统运行控制模型特性的电参量及涉电辅助参量数据。

1. 电源侧：建模数据应包括各类电站小时级或15分钟级的有功功率、无功功率、相角，母线电压、各出线电流及其相位值，各开关及刀闸变位信息等遥控、遥测、遥信、遥调等信息，机端调速器信息及稳控、继保等二次设备信息等。除此之外，对于水电，还应包括实时的机组运行区间、坝前水位信息、发电流量、下泄流量及防洪控制要求等；对于火电，还应包括实时的储煤量、煤耗量，及其他附属设施运行参数。
2. 电网侧：对于输电线路，建模数据应包括线路实时潮流情况，电压损耗情况，线路两侧无功补偿情况，两侧开关及刀闸变位信息，线路相关继电保护装置及其他安全自动装置的实时运行工况，防火、防雷监测系统运行状况等；对于配电网，建模数据应涵盖多种配电业务，涉及配电设备、配网运行状态数据；对于变电站，建模数据应包括全站电参量，类如全站各设备电流、电压、相位、潮流情况，主变运行状态，类如主变油温、频谱分析结果、瓦斯气体含量等，全站设备“四遥”信息及开关、刀闸变位信息，继保、稳控的二次设备实时运行工况及相关报文信息。
3. 负荷侧：建模数据应涵盖用户及相关附属设施全部电参量及相关涉电参量，包括负荷侧变压器和馈线潮流、电压、电流、相位信息等，及其相关附属二次设备运行工况信息，用户相关计量表计运行数据及其工况信息等；其中用电数据还应包含超短期负荷及分布式可再生能源预测数据，帮助用户了解未来发用电情况。
4. 储能系统：建模数据应包括设备类型、设备容量、充放电效率、连续放电时间及其实时工况信息，量化储能系统使用寿命及风险预警信息。
	* 1. 行为决策模型的数据要求

行为决策模型数据包括描述源-网、网-网、网-荷、源-荷、网-储、荷-储、源-储等场景协调过程行为相关参量。

1. 源-网协调场景，主要行为决策模式为调控一体化，主要涉及影响电网有功、无功平衡及电源涉网协调等行为决策的量化指标，应包括弃风、弃光、弃水电量，年最大利用小时数，旋转备用容量，机组调峰能力，低电压穿越能力，高电压穿越能力，稳控切机量及切机顺序敏感度分析等。
2. 网-网协调场景，主要行为决策模式为跨区域、跨省级、跨地市等不同电网间的协调调度，主要涉及开机方式，旋备支援，阻塞容量，电力、电量实时交换、冲抵等市场化行为。
3. 网-荷协调场景，主要行为决策模式为营配贯通，主要涉及用户用电量及其用电行为特征，市场化交易行为及其他用户涉网服务信息等。
4. 源-荷协调场景，主要行为决策模式为电力交易协议、大用户直购电及其他涉电交易行为。
5. 网-储协调场景，主要行为决策模式为储能系统参与电网侧辅助服务市场的相关涉网行为，包括参与调峰，提供事故备用等辅助服务市场行为。
6. 荷-储协调场景，主要行为决策模式为储能系统参与用户侧辅助服务市场的相关涉网行为，包括提高电能质量，获取峰谷电价差等。
7. 源-储协调场景，主要行为决策模式为储能系统参与电源侧辅助服务市场的相关涉网行为，包括平抑新能源波动功率、提高电源出力稳定性，获取峰谷电价差等。
8. 数据预处理技术及要求
	1. 数据预处理基本技术
		1. 数据清理

实际采集的数据一般是不完整、有噪声的。相应地，数据清理的基本技术包括：

1. 处理缺失值。可采用忽略属性缺失的数据对象、人工填写缺失值、使用全局常量填充缺失值、使用属性值的中心度量填充缺失值、使用最可能的值填充缺失值等方法。
2. 处理噪声值。可采用分箱平滑、回归平滑、离群点分析等方法。
	* 1. 数据集成

对于多个数据来源，需要采用数据集成技术合并各数据来源，形成单一数据集。数据集成的基本技术包括：

1. 实体识别与合并。识别代表同一实体的多个数据对象，将其属性值进行合并，解决冲突问题。
2. 冗余与相关分析。若一个属性能由另一个或另一组属性导出，则可能为冗余属性。检测冗余的方法为相关分析。包括用于标称数据的$χ^{2}$相关检验、用于数值数据的相关系数、协方差分析。
	* 1. 数据归约

针对大规模数据集，需要采用数据归约技术降低数据规模并尽可能保持原始数据的完整性，从而减少数据存储压力，提高数据加载、处理与分析的速度。数据归约的基本技术包括：

1. 维度归约。删除不相关、弱相关或冗余属性，减少所考虑的属性的个数。
2. 数量归约。用替代的、较小的数据表示形式替换原数据，包括参数方法与非参数方法。参数方法使用模型估计数据，只存放模型参数与离群点；非参数方法包括直方图、聚类、抽样等。
3. 数据压缩。使用压缩算法削减数据规模，原数据可以从压缩后的数据中无损重构或近似重构。包括小波变换、主成分分析等。
	* 1. 数据转换

针对异源数据，通常需要通过数据转换得到适合数据建模的格式。数据转换的基本技术包括：

1. 属性构造：由给定的属性构造新的属性并添加到属性集中，以帮助数据建模使用。
2. 规范化：把属性数据按比例缩放，使之落入一个特定的区间。
3. 离散化：把数值属性的原始值用区间标签或概念标签替换。
	1. 预处理后的数据质量要求
		1. 错误率

数据值应当与期望的真实值一致，对于经过预处理后的数据集，分数值数据与标称数据分别规定错误率指标如下：

用公式（1）计算出数值数据错误率指标。

 $相对均方根误差=\frac{\sqrt{\frac{\sum\_{i=1}^{N}\left(\hat{x}-x\right)^{2}}{N}}}{\frac{\sum\_{i=1}^{N}\hat{x}}{N}}×100\%$ （1）

式中：

$\hat{x}$为数据值；

$x$为真实值；

$N$为数据集样本总数量。

用公式（2）计算出标称数据错误率指标。

 $标称数据错误率=\frac{F}{N}×100\%$ （2）

式中：

$F$为标称属性错误的样本数量；

$N$为数据集样本总数量。

引起数据不准确的原因包括：数据采样与传输过程中产生误差、输入字段格式错误、实际场景中用户提交数据出错、数据转换中的计算误差叠加。

经过预处理后的数据应满足规定的错误率限制，才能进一步应用于数据-机理融合建模任务。按照错误率指标度量预处理后数据集的准确性，可划分数据集准确性等级：

A类数据集应满足错误率低于1%；

B类数据集应满足错误率落于1%到5%区间；

C类数据集应满足错误率落于5%到10%区间。

* + 1. 完整性

数据应当完整，不应明显缺少实体数据或缺少属性值。引起数据不完整的基本因素有：

1. 数据不可用。
2. 删除不一致数据。
3. 删除无关数据。

针对以上引起数据不完整的基本因素，经过预处理后的数据集应确保完整性要求，包括：

1. 实体完整性：预处理后数据集应该覆盖绝大部分实体；A类数据集应满足实体覆盖率大于95%；B类数据集应满足实体覆盖率落于90%到95%区间；C类数据集应满足实体覆盖率落于80%到90%区间。
2. 属性完整性：全部数据对象不应缺少属性值。
	* 1. 一致性

对于由多种数据来源聚合而成的数据集，经预处理后不应出现多个数据对象对应一个数据实体的冲突。应包含以下几个方面：

1. 数据实体一致性。
2. 属性一致性。
3. 存储格式一致性。
4. 归约格式一致性。
5. 数据-机理融合建模信息数据接口建议

从“源-网-荷-储”架构出发，以行为决策模型数据接口为例，融合建模数据接口建议参见表A.1，具体可根据实际情况进行调整。

表A.1　融合建模信息数据接口建议表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **对象名称** | **属性名称** | **描述** | **类型** | **长度（字节）** | **说明** |
| 发电机 | GEN\_ID | 机组ID | unsigned int | 4 |  |
| GEN\_NAME | 机组名称 | 64\*char | 64\*1 |  |
| GEN\_TYPE | 机组类型 | unsigned int | 4 | 1：常规火电；2：可启停火电；3：燃机；4：核电；5：水电；6：抽蓄；7：热电；8：联合循环燃机；9：风电；10：光伏 |
| GEN\_BUS | 所在节点 | 64\*char | 64\*1 |  |
| GEN\_PLANT | 所属电厂 | 64\*char | 64\*1 |  |
| PMAX | 机组容量 | float | 4 | 单位：MW |
| PMIN | 最低出力 | float | 4 | 单位：MW |
| COST\_S | 启停费用 | float | 4 | 单位：万元 |
| COST\_FIXED | 固定运行费用 | float | 4 | 单位：万元/年，包括保险费、燃料储备费用、运行维护费(固定部分)、工资及行政开支等 |
| COST\_VMAX | 满出力时可变运行费 | float | 4 | 单位：元/kWh，，包括燃料费、维护费(可变部分)、财务及税金等 |
| COST\_VMIN | 最低出力时可变运行费 | float | 4 | 单位：元/kWh |
| CRATE\_OMAX | 最高出力时煤耗率或气耗率 | float | 4 | 单位：对燃煤机组kg/kWh；对燃气机组m3/kWh |
| CRATE\_OMIN | 最低出力时煤耗率或气耗率 | float | 4 | 对燃煤机组kg/kWh；对燃气机组m3/kWh |
| TYPE\_AGC | 是否为AGC机组 | bool | 1 bit | 1表示AGC机组，0表示非AGC机组 |
| RAMP\_AGC | 机组调频调节速率 | float | 4 | 单位：MW/min |
| RAMP\_15 | 机组负荷跟踪调节速率 | float | 4 | 单位：MW/15min |
| RAMP\_UP | 机组爬坡UP调节速率 | float | 4 | 单位：MW/h |
| RAMP\_DOWN | 机组爬坡DOWN调节速率 | float | 4 | 单位：MW/h |
| Num\_S | 日内最大启停次数 | unsigned int | 4 | 单位：次 |
| Time\_UP | 最小开机时间 | float | 4 | 单位：h |
| Time\_DOWN | 最小停机时间 | float | 4 | 单位：h |
| POWER | 历史发电信息 | float\*时段 | 4\*时段 | 可根据风电、光伏历史出力信息进行时序预测 |
| 节点 | BUS\_ID | 节点ID | unsigned int | 4 |  |
| BUS\_NAME | 节点名称 | 64\*char | 64\*1 |  |
| BUS\_AREA | 所属区域 | unsigned int | 4 |  |
| BUS\_TYPE | 节点类型 | unsigned int | 4 |  |
| MBASE | 基准容量 | float | 4 | 单位：MVA |
| 线路 | F\_BUS | 起始节点 | unsigned int | 4 |  |
| T\_BUS | 终止节点 | unsigned int | 4 |  |
| BR\_R | 线路电阻 | float | 4 | p.u. |
| BR\_X | 线路电抗 | float | 4 | p.u. |
| BR\_B | 线路电纳 | float | 4 | p.u. |
| BR\_M | 传输容量 | float | 4 | 单位：MW |
| BR\_TYPE | 线路类型 | unsigned int | 4 | 1：1000kV线路；2：1000kV/500kV变压器；3：1000kV母联开关；4：500kV线路；5：500kV/220kV变压器；6：500kV机端变压器；7：500kV母联开关；8：220kV线路；9：220kV机端变压器；10：220kV母联开关；20：直流线路 |
| BR\_LENGTH | 线路长度 | float | 4 | 单位：km |
| BR\_STATUS | 线路状态 | float | 4 | 1表示在运行，0表示在检修 |
| 负荷 | LOAD\_BUS | 所在节点 | 64\*char | 64\*1 |  |
| PD | 历史用电信息（有功功率） | float | 4 | 单位：kW |
| PRICE | 历史电价信息 | float | 4 | 单位：元/kWh |
| 储能 | STOR\_ID | 储能ID | unsigned int | 4 |  |
| STOR\_NAME | 储能电站名称 | 64\*char | 64\*1 |  |
| STOR\_TYPE | 储能类型 | unsigned int | 4 | 1：锂电池；2：铅酸电池；3：钒液流电池；4：液流电池；5：钠硫电池 |
| STOR\_BUS | 所在节点 | 64\*char | 64\*1 |  |
| STOR\_M | 装机容量 | float | 4 | 单位：MW |
| STOR\_CF | 固定运行费用 | float | 4 | 单位：万元/年 |
| STOR\_CV | 可变运行费用 | float | 4 | 万元/kWh |
| STOR\_EC | 充电效率 | float | 4 |  |
| STOR\_ED | 放电效率 | float | 4 |  |
| STOR\_TIME | 持续放电时长 | float | 4 | 单位：h |
| STOR\_CYL | 平衡周期 | float | 4 | 单位：h |