基于电信数据的高速场景新能源车充电需求刻画算法

**北京大学深圳研究生院**

**华为科技有限公司**

**2023年03月**

目录

[一、 项目背景【余庆】 1](#_Toc17766)

[二、 技术路线【余庆】 2](#_Toc8718)

[2.1 整体技术路线 2](#_Toc23007)

[三、 高速公路拓扑网络建模【徐剑】 3](#_Toc21568)

[3.1 整体技术路线 3](#_Toc7695)

[3.2 路网预处理 4](#_Toc31358)

[3.2.1 路网简化 4](#_Toc18830)

[3.2.2 双向路段生成 5](#_Toc14712)

[3.2.3 节点近邻匹配 5](#_Toc29661)

[3.2.4 路网调整 6](#_Toc500)

[3.2.5 双向道路面生成 8](#_Toc19660)

[3.3 路网建模与出行路径 9](#_Toc17949)

[3.3.1 拓扑网络建模 9](#_Toc24844)

[3.3.2 OD出行路径获取 9](#_Toc32331)

[3.3.3 节点最近收费站匹配 10](#_Toc5206)

[四、 高速公路出行OD推测 11](#_Toc8462)

[4.1 模块功能介绍 11](#_Toc1487)

[4.2 OD推测算法1：优化模型 11](#_Toc4220)

[4.2.1 MR热力流量提取 12](#_Toc30870)

[4.2.2 优化模型构建 12](#_Toc20247)

[4.3 OD推测算法2：百度迁徙 15](#_Toc27029)

[4.3.1 百度迁徙数据获取 16](#_Toc8394)

[4.3.2 收费站出行OD推断 16](#_Toc1295)

[4.4 OD结果调整 17](#_Toc8387)

[4.4.1 天级到小时级调整（韶关小时级数据 17](#_Toc395)

[4.4.2 出行距离调整 17](#_Toc23356)

[五、 车辆时空轨迹与行为仿真【刘思远】 18](#_Toc21244)

[5.1 ABM仿真技术介绍 18](#_Toc12913)

[5.2 ABM仿真代理模型 19](#_Toc98)

[5.2.1 车辆代理 19](#_Toc20560)

[5.2.2 充电站代理 19](#_Toc30516)

[5.2.3 代理设计 20](#_Toc12970)

[5.3 仿真模型参数与场景设置 22](#_Toc21330)

[5.3.1 仿真模型参数： 22](#_Toc3566)

[22](#_Toc1873)

[5.3.2 场景设置： 22](#_Toc25266)

[5.3.3 模型优势： 23](#_Toc28767)

[六、 服务区充电需求预测评价与场站配置建议【李家兴】 24](#_Toc17563)

[6.1 上游路段MR与充电车辆数比较结果 24](#_Toc31249)

[6.2 仿真的结果比较分析 25](#_Toc1373)

[6.3 选址参考指标 28](#_Toc23003)

# 项目背景【余庆】

# 技术路线【余庆】

## 整体技术路线

空间数据是描述自然地理空间和人类活动空间所包含的人、物体、事件的信息[1]，在此基础上拓展人工智能技术，在算法、数据、计算力等方面持续创新，实现空间数据智能技术。空间数据智能实现的关键技术包括空间感知、空间数据存储、空间/时空数据挖掘、边缘计算与雾计算等多个方面。

# 高速公路拓扑网络建模【徐剑】

## 整体技术路线

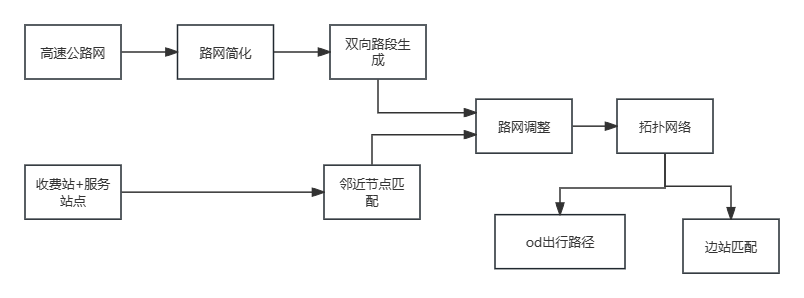
在高速公路拓扑网络建模这一模块中，旨在处理和分析高速公路网络的数据。首先，从高速公路网络中提取边界的几何信息和唯一标识符（ID），同时创建反向边以确保建立完整的道路网络。为了确保站点与道路边界的准确匹配，进行了坐标投影和偏移操作，并利用KDTree算法找到每个站点最近的道路边界。

随后，进行了网络结构的一系列调整，包括边界的切割和新线段的添加，为每个道路的起点和终点分配唯一的ID。然后，边界的坐标信息进行了转换，生成多边形对象，这些多边形对象代表了道路的形状和宽度，以确保与人口热力数据的精确匹配。

通过建立拓扑网络，为了路径计算，将边的权重设置为其长度，以构建出行的起点和终点的OD（Origin-Destination）组合。最后，执行最短路径计算，获取路段信息，计算累积总长度，并将结果保存以供后续分析使用。这一系列操作旨在更好地理解和利用高速公路网络的信息。

该技术路线的目的在于对高速公路网络进行简化和分析，以实现更高效的路网数据表示和路径计算，从而为后续仿真研究提供支持和帮助。

整体技术流程图：



整体流程输入：

|  |  |  |
| --- | --- | --- |
| 输入（Input） | 字段（Fields） | 解释 |
| Edge | geometry | 边的地理信息 |
|  | edge\_id | 边的编号 |
| station | id | 收费站的ID |
|  | geometry | 收费站的地理信息 |

整体流程输出：

|  |  |  |
| --- | --- | --- |
| 输出（Output） | 字段（Fields） | 解释 |
| od | station\_id\_x | od出行路径的起点收费站ID |
|  | station\_id\_y | od出行路径的终点收费站ID |
|  | length | 出行路径的距离 |
| edge\_to\_charge | edge\_id | 边的ID编号 |
|  | charge\_node\_id | 充电站的ID编号 |
|  | charge\_node\_distance | 边到最近充电站节点的距离 |
|  | 场站名称 | 充电站的名称 |
|  | 充电车位数 | 充电站的充电车位数 |

## 路网预处理

### 路网简化

路网简化是指对高速公路网的矢量数据进行简化，未简化的高速公路网将包含 较多车道、立交、交叉口等冗余矢量信息，简化后的高速公路网将只包括双向车道、 服务区等关键信息。

简化路网使用ArcGIS软件操作的打断相交线进行操作。在 ArcGIS 中，打断相 交线是指对两条或多条线进行分割，使它们在相交点处打断并保留其几何关系。此 操作有助于确保线要素在相交处保持连续，而不会形成多余的节点。

打断相交线的方法如下：

1. 编辑模式: 打开 ArcMap 或 ArcGIS Pro，进入编辑模式。

2. 选择工具： 选择编辑工具栏，编辑器—更多编辑工具—拓扑，添加“拓扑” 工 具条，点击“拓扑”工具条上的“打断相交线”工具。

3. 选择要素: 选择需要分割的线要素图层。

4. 执行段相交线: 在弹出的“打断相交线”界面，输入拓扑容差值后，点击“确定”， 高速公路网的拓扑容差设置为0.0003。

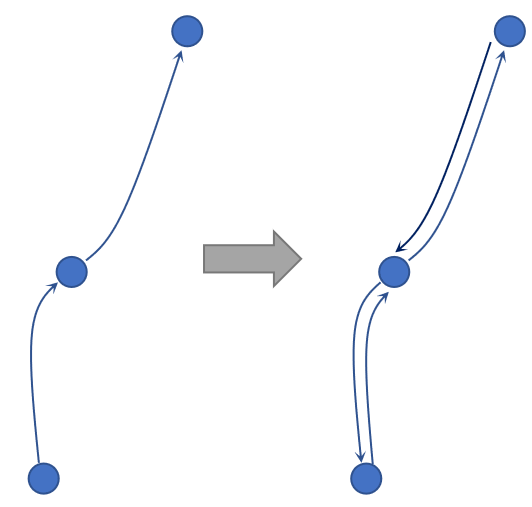
5. 保存编辑: 当打断相交线完毕后，保存编辑并退出编辑模式。

### 双向路段生成

生成双向路段的目的是为了准确表示道路的双向通行情况。通过复制和反向处 理几何信息，我们可以创建与原始边相对应的反向边，以完整地描述道路的双向行 驶能力。将正向边和反向边合并到一个DataFrame中，有助于方便后续的路网分析 和路径计算。

具体步骤如下：

从GeoJSON文件中读取了高速公路网络信息，提取了几何信息并生成了边 ID。然后，通过复制并逆序处理几何信息，生成了反向边。最后，将正向边和 反向边合并到一个DataFrame中。



### 节点近邻匹配

节点近邻匹配的技术目的在于实现站点与道路边的精确匹配，以便后续的路网 分析和应用。通过将站点和边的坐标系投影到适用于广东省的坐标系，对边进行偏 移，以区分双向道路，确保站点正确对应相应的路段。利用KDTree算法，能够有 效地找到每个站点最接近的道路边，从而建立站点和道路边之间的关联关系，并生 成包含站点ID和最近道路边ID的新数据表。

具体步骤如下：

首先，通过将站点（station）和边（edge\_offset，是道路边的副本）的坐标系设置为地理坐标系EPSG 4326，然后使用to\_crs函数将它们投影到广东省适用的坐标系EPSG 4525（CGCS2000 / 3-degree Gauss Kruger zone 37），实现了坐标系的调整。这一步骤的目的在于确保数据的一致性和适应性。

接下来，为了确保道路边在不同方向上都能正确匹配，edge\_offset中的每一条边的几何信息都经过偏移处理。这一偏移操作旨在将双向道路分隔开，以确保每个站点（station）能够正确匹配到相应的道路方向。如果不进行这一分离，一个站点可能会对应到两条边，但实际上应该只对应到一条。

随后，我们调用了TransBigData中的ckdnearest\_line函数。该函数底层使用KDTree算法来查找每个站点离其最近的道路边。有关该函数的详细说明可在https://transbigdata.readthedocs.io/zh-cn/latest/gisprocess.html#transbigdata.ckdnearest\_point找到。这一步骤至关重要，因为它建立了站点与道路边之间的关联关系，并生成了一个新的数据表，其中包含站点ID和最近道路边的ID。

最终，经过节点近邻匹配的操作，生成的新表格被妥善保存。该表格包括站点的经度（lon）和纬度（lat），对应每个站点的唯一编号（id），以及站点的名称等信息。这个表格可供后续的分析和应用使用，有助于更深入地研究和规划交通网络，以满足特定需求和决策制定。这一系列步骤为数据的准确性和可用性提供了坚实的基础。

### 路网调整

鉴于原始数据存在许多站点信息的缺失，旨在提升网络数据的完整性，需要将 额外抓取的站点数据整合到原始网络中。

具体步骤如下：

首先，针对每个节点与其最近边的匹配结果，提取站点ID、边ID以及相关几何信息。

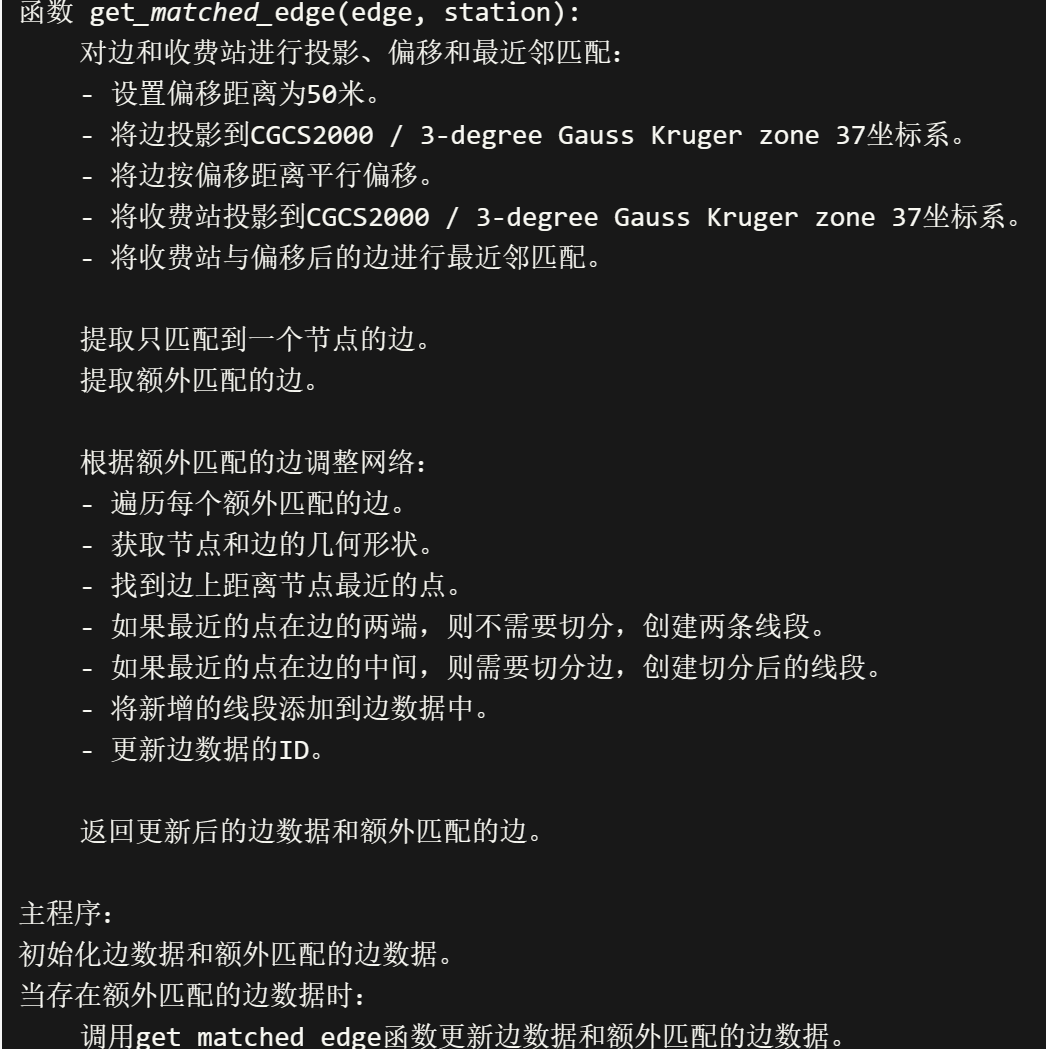
接着，利用节点和边的信息，计算边上距离节点最近的点以及该点到边的投影距离。这一过程的核心代码中，projectdist表示边的起点。

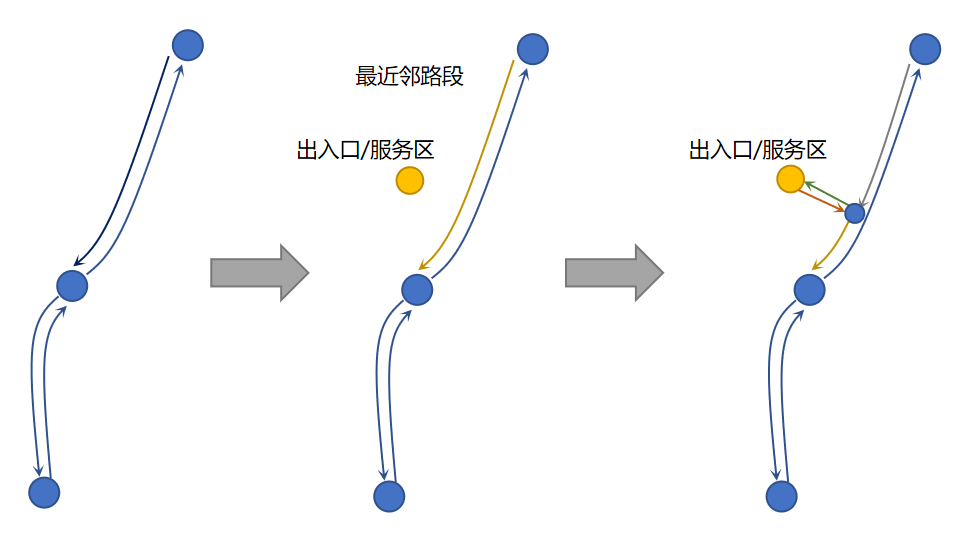
然后，判断最近点是否位于边的两端。若是，直接将端点与节点之间的线段添加到网络中。如果最近点位于边的中间位置，需要对原始边进行分割，并添加端点到节点的线段。

接下来，将分割后的边段以及连接线段添加到新的边列表中。将新的边列表与原始边数据合并，从而得到经过调整的网络。为确保每条边都具有唯一标识，重新生成边的ID。

进一步，从“geometry”字段中提取边表中每条边的起点和终点的经纬度，即OD 信息。最后，为所有起点和终点分配唯一的ID编号，并将这些信息合并到边表中。这一系列步骤将有助于完善网络数据，确保站点与道路边的关联，以支持后续的网络分析和应用。

算法流程图如下：





### 双向道路面生成

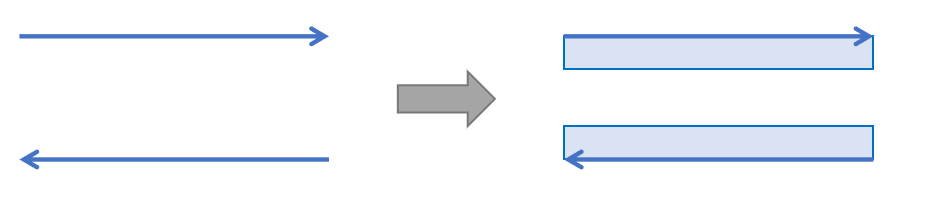
将道路线转化为面，以实现更精确的道路与人口热力数据的匹配，从而提升数 据质量和关联的准确性。这一转化过程丰富了数据维度，加强了道路特征的完整性， 为进一步的分析和应用提供了更为丰富的信息。

具体步骤如下：

首先，将原始边的坐标系转换为适用的投影坐标系（EPSG:4326）。接着，添 加一个名为“length”的列，用于储存每条边的长度信息。

然后，将边的几何信息转 换为广东省适用的CGCS2000 / 3-degree Gauss Kruger zone 37投影坐标系 （EPSG:4525）。通过使用shapely.geometry.Polygon创 建一个新的多边形对象，将 包括边的坐标和右侧一定距离的平移坐标。这一步骤旨 在生成道路的面状表示，以 丰富道路特征。

最后，将生成的面对象的几何信息重新转换回原始的EPSG:4326坐 标系，以确 保数据的一致性和适应性。这一系列操作有助于将道路线转化为面，从 而提供更 精确和详尽的道路信息，以支持进一步的分析和应用。



## 路网建模与出行路径

### 拓扑网络建模

将数据进行转化，以构建完整的网络拓扑结构，以便于执行最短路径计算以满 足每个出发地（OD）对之间的需求。这一操作丰富了数据处理的层次，确保网络的 结构和关系得以明确，为有效地计算最短路径提供了基础。

具体步骤如下：

使用geopandas库读取保存的边，节点，收费站，道路面数据，使用network 库 创 建有向图，设置权重位边的长度。

### OD出行路径获取

提取出OD数据，以备后续进行仿真实验和模拟研究。这一步骤不仅有助于数 据精炼和集中，还为进行模拟实验提供了必要的输入。

具体步骤如下：

首先，根据收费站节点对应网络节点的编号数据以及OD表，创建了包括所有 可能的出行OD组合的数据集。

随后，通过利用networkx库的shortest\_path方法， 计算了每个出行OD组合 的最短路径，并获取了路径上的节点序列。为了获得出行路段信息，定义了一个 名为get\_path\_gdf的函数。这个函数的主要功能是将节点序 列转化为相应的边序 列，并将其与道路边的数据进行匹配，进而计算出行路径的累计长度。

另外，还定义了一个名为get\_path\_dis\_table的函数，用于获取每个OD出行所 经过的路段信息。这个函数将路径信息转化为路段信息，并为每个OD添加了其起 点和终点的收费站编号。

通过以上步骤，成功计算并获得了每个OD出行所经过的路径信息，包括累计 总长度，这些信息已被妥善保存以供后续分析和研究之用。

### 节点最近收费站匹配

根据给定的高速公路网络图，计算每个节点距离最近的收费站节点，并将相关信息存储到CSV文件中。

具体步骤如下：

首先，从文件中读取包含边、节点和收费站节点对应关系的数据。然后，创建 一个有向图对象，并向图中添加节点。接下来，将边添加到图中，以构建完整的网 络拓扑。

定义了一个名为get\_nearest\_charge\_station的函数，用于找到每个节点距离 最 近的收费站节点。在给定节点的半径范围内，创建一个子图。然后，遍历子图中 的节点，并计算每个节点与收费站节点之间的最短路径距离。函数返回距离最近的 收费站节点的ID和距离。接下来，对每个节点调用get\_nearest\_charge\_station函数， 以找到每个节点距离最近的收费站节点。

创建一个包含节点ID的DataFrame对象，用于存储节点的信息。使用并行计算 （parallel\_apply）遍历每个节点ID，并调用get\_nearest\_charge\_station函数，得到每 个节点距离最近的收费站节点ID和距离。将这些最近的收费站节点ID和距离添加 到节点DataFrame中。

将边DataFrame与节点DataFrame合并，以获取边与最近收费站节点的对应关 系。将合并后的数据与收费站数据进行合并，获取边与最近收费站节点的详细信息。

按照边ID进行排序，以便更好地组织数据。

# 高速公路出行OD推测

## 模块功能介绍

该模块的主要功能是通过高速网络拓扑建模之后生成的路网的拓扑数据，与MR热力数据进行叠加，通过优化模型对每个OD对（O、D为高速公路出入口）之间的流量进行推测与验证，进而推断OD分布的概率。同时考虑到了不同时段，不同出行距离和出行意愿的影响，并根据相应的数据对新能源汽车的小时级OD分布概率进行了修正。

整体流程输入：

|  |  |  |
| --- | --- | --- |
| 输入文件（Input） | 字段（Fields） | 字段解释 |
| mr\_edge\_count\_opt.csv | edge\_id | 高速公路路段编号 |
| MR\_count\_median | MR数据与高速公路叠加后得到的路段热力数据的中值 |
| date | MR数据对应的日期 |
| od\_dis\_table\_opt.csv | station\_id\_x | 作为O点收费站的ID |
| station\_id\_y | 作为D点收费站的ID |
| edge\_id | 收费站之间最短路径经过的所有路段的编号 |
| edge\_to\_charge.csv | edge\_id | 建设有充电站，需要进行充电判断的路段编号 |

整体流程输出：

|  |  |  |
| --- | --- | --- |
| 输出（Output） | 字段（Fields） | 字段解释 |
| OD\_flow\_{date}.csv | O | 作为O点收费站的ID |
| D | 作为D点收费站的ID |
| count/prop | 根据OD推测得到的收费站之间的流量/概率 |
| date | 对应日期 |

## OD推测算法1：优化模型

大规模的矩阵求解对于内存和cpu都有较高的要求，并且通常需要极大的的时间成本，直接求解困难较大。因此，该OD推测算法的核心思路是将大规模的OD-edge矩阵拆分成若干个小矩阵，并利用最小二乘和约束最小二乘法进行求解，运用到的数据主要是表征每个OD对（O、D为高速公路出入口）需要经过的路径的OD-edge矩阵，以及表征每条路段上车流量的的MR热力数据的中值MR\_count\_median。

具体的步骤包括：

（1）对MR热力数据提取到高速公路路段上，得到每条路段的的流量。

（2）在优化模型中输入OD—edge的稀疏矩阵和路段edge的流量表。

（3）将矩阵进行分割，然后计算分割后数据的比例，删除数据矩阵和流量数据中的无效信息。

（4）利用带有正则项的最小二乘法进行求解，迭代删除OD对和edge，直到满足相对误差阈值sigma或OD数小于edge数。

（5）利用第一次最小二乘法处理后的数据，使用约束最小二乘法进行求解，得到每对OD之间的流量。

### MR热力流量提取

在路网拓扑建模的基础上，利用geopandas库将每个路段与MR热力数据的网格在空间上叠加，得到每个路段上对应的多个热力值，对所有的热力值求中位数来代表每个路段上的车流量。具体步骤如下：

1、读入MR热力数据和高速公路路网的拓扑数据并进行预处理，利用geopandas库将栅格的经纬度坐标转化为点要素信息，以便进行后续分析统计。

2、将点要素与高速公路面数据在空间上进行叠加，分别统计每条edge上的点的数量MR\_count\_grids，MR的总和MR\_count\_sum以及中值MR\_count\_median。

3、读入之前处理得到的记录OD对途径路径的文件od\_dis\_table.csv，筛选出长度大于1000m的路径并记录编号，根据编号对之前得到路段edge的热力数据进行筛选，去除不必要的路段数据，导出具有大于1000m的路段的信息的文件od\_dis\_table\_opt.csv和MR\_edge\_count\_opt.csv。

### 优化模型构建

在data\_processing.py程序中，程序首先实现数据的预处理工作，输入数据文件为od\_dis\_table\_opt.csv和mr\_edge\_count\_opt.csv。由于部分路段在某些天可能存在缺失MR热力值的问题，因此首先对od\_dis\_table文件和mr\_edge\_count文件中的路径edge\_id求交集，并删除存在缺失值的路段。进一步根据OD及其经过的edge构建矩阵，矩阵每行代表一条路段edge，矩阵每列代表一个OD对，进一步利用 库的 sparse函数，将该矩阵转化为稀疏矩阵，并保存为 .npz。将筛选后的mr\_edge\_count按照edge\_id进行排序，并导出保存为 .csv。

|  |
| --- |
| **# 程序说明**  # 运行此脚本的文件夹结构如下：  # ----/scaling\_factor\_estimation  # |----/scripts  # | |----main\_scaling\_factor.py <-- 本文件  # |----bw\_data\_matrix.npz <-- 来自 preprocessing.py 的结果  # |----scenario\_table.csv <-- 来自 preprocessing.py 的结果  # |----/results |
| **#数据输入**  # 程序输入 1：OD\_edge\_matrix.npz 是一个具有 n 行和 r 列的稀疏矩阵。其中，n 表示高速公路收费站OD对的数量，r表示高速公路路段的数量。  # 程序输入 2：edge\_flow.csv 是一个具有 r 行和 1 列的数组。  # 程序输出：'ODresults\_{train\_line}\_{start\_day}\_{start\_hour}\_{start\_minute}.csv' 包含所有 OD对的流量数据，排序方式与OD\_edge\_matrix.npz矩阵相同。 |
| **# 参数设置**  # 设置容差参数sigma。如果相对误差小于 'sigma'，则算法终止。  # sigma = 0.1 # 浮点数 sigma >= 0：当 sigma 增加时，程序运行时间减少  # num\_edge\_delete = 5 # 整数 num\_edge\_delete >= 2：当 num\_edge\_delete 增加时，程序运行时间增加  # num\_od\_delete = 5 # 整数 num\_od\_delete >= 1：当 num\_od\_delete 增加时，程序运行时间增加  # max\_iter\_num = 100 # 整数 max\_iter\_num >= 5：当 max\_iter\_num 增加时，程序运行时间增加，并提高精度 |

在main\_FlowEstimation程序中，主要实现了利用最小二乘法求解OD流量。具体的步骤如下：

1. 导入所需的库和模块，包括：time、pandas、numpy、scipy、argparse等。
2. 定义一个FlowEstimator类，该类包含了一些参数和方法。首先在类的初始化方法中，设置了一些默认参数，包括sigma（相对误差的阈值）、num\_scenario\_delete（要删除的场景数）、num\_user\_delete（要删除的用户数）和max\_iter\_num（约束最小二乘法的最大迭代次数）。
3. FlowEstimator类的fitting方法是主要的计算逻辑。它接受两个输入参数：matrix\_data和edge\_data,分别读取输入的数据文件OD\_data\_matrix.npz和edge\_flow.csv，这些数据分别代表OD和edge的稀疏矩阵数据和每个路段的车流量数据。同时该方法还包括三个默认参数division\_base（分割后矩阵的尺寸）、train\_line（选择需要训练的数据，这里是日期），mode（决定是否要保存输出结果，当等于0时则输出结果）。
4. 对数据进行预处理，包括加载稀疏矩阵和路段的车流量数据，提取相关数据并对其进行处理：

|  |
| --- |
| 根据train\_line参数选择路段的车流量数据需要处理的数据edge\_ |
| 转换稀疏矩阵的数据类型并对其进行转置，得到处理后的矩阵od\_matrix\_ |
| 计算od\_matrix\_中非零元素的和od\_sum\_total。 |
| 根据给定的division\_base参数将矩阵od\_matrix\_切分成若干小矩阵，计算迭代次数 |
| **对于每次迭代：** |
| 根据division\_base切分od\_matrix\_，计算切分后的od\_matrix的比例od\_ratio。 |
| 根据od\_ratio计算edge\_的比例edge |
| **删除s\_matrix中信息较少的OD和路径**： |
| 计算每列的和od\_matrix\_sum\_index，并找到大于0.5的列索引mid1\_index |
| 根据mid1\_index筛选出s\_matrix的子矩阵od\_matrix\_p\_resolve1 |
| 计算od\_matrix\_p\_resolve1的每行的和，并找到大于0.5的行索引mid2\_index |
| 根据mid2\_index筛选出od\_matrix\_p\_resolve1的子矩阵od\_matrix\_p\_resolve2 |
| 根据mid2\_index筛选出edge的子矩阵edge\_p\_resolve2 |
| 打印被删除的OD和edge的信息 |

1. 根据Tikhonov正则化方法来对数据进行处理，去除不必要的OD和edge。具体是将矩阵正则化后使用一般最小二乘法进行计算，并将预测的结果与真实结果进行对比，计算其绝对平均误差。按照信息量从少到多的顺序来对OD和路径进行删除，直到满足绝对平均误差小于给定的相对误差sigma或OD数小于路段数。

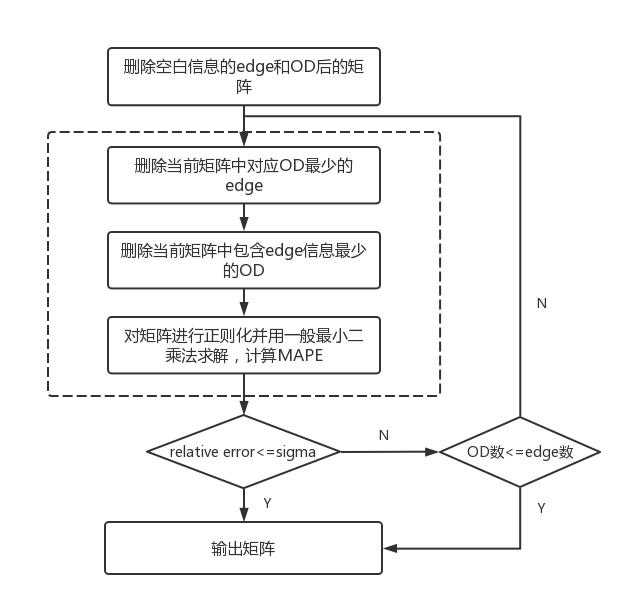


图 1

1. 将满足求解条件的数据利用scipy库的线性求解库中的scipy.optimize.lsq\_linear命令来实现约束最小二乘法的求解，设置每个OD对的流量的上下限和模型的迭代次数max\_iter\_num,以保证模型收敛并且每个OD对的流量数据在合理的范围内，求解得到满足约束的每个OD对的流量计算结果。
2. 导出计算结果到文件，文件中的结果数值的排列顺序与输入的稀疏矩阵中的排列顺序一致：ODresult\_{train\_line}{division\_base}{start\_day}{start\_hour}{start\_minute}.csv。
3. 使用argparse模块解析命令行参数，并根据参数运行FlowEstimator类的fitting方法。
4. 打印程序执行完成的提示信息,输出程序运行的时间。

最后通过match.py程序对输出的结果数据进行处理，利用pandas库的merge命令为结果添加O、D以及date信息，输出包含OD和日期信息的结果文件。

## OD推测算法2：百度迁徙

该算法主要基于高速公路出入口的MR热力数据和广东省内的地级市之间的百度迁徙数据，通过出入口的MR热力数据计算出每两个城市之间各高速公路出入口的权重，并将地级市之间的人口流动的规模指数分配到每个OD上，计算出每个OD的分布概率。该算法较为简单，计算效率高，同时通过引入跨省的百度迁徙指数，可以对跨省的交通流量进行估计。

### 百度迁徙数据获取

百度迁徙数据能够较为直观、全面地描述和表现大范围内人口流动的整体情况以及城际交通运输需求的时空分布特征。对广东省内各市的百度迁徙数据进行爬取，得到各市之间的迁徙规模指数，其中迁入指数表示在指定日期迁入某一城市i的客流规模，迁出指数表示在指定日期迁出某一城市j的客流规模。

通过将百度迁徙得到的广东省内各城市之间不同日期的城际客流规模数据可视化，可以初步得到人口在研究时期内的大致流向。

### 收费站出行OD推断

收费站出行OD推断主要依据地级市之间的百度迁徙数据，结合MR热力数据来进行计算。

1、首先需要将广东省各个高速公路的出入口分配到地级市。利用python的Geopandas库，通过将高速公路出入口点的坐标与各市的面域数据叠加，得到并统计各个出入口所在的地级市信息。依据高速公路的出入口数据生成所有可能的OD对，并将每个OD对的OD对应到地级市层级。

2、基于高速公路的出入口数据，所有可能的OD（Origin-Destination）对将被生成。每个OD对将映射到地级市层级，以建立高速公路交通流动的起点和终点。同时为了确定每个OD对之间的流量，华为云瞰数据中的MR热力数据将被叠加在各高速公路出入口上，得到出口O\_count和入口D\_count的热力数据，并根据这些热力数据来计算每个高速公路入口O和高速公路出口D的概率prop，将其相乘可以得到每个OD的分布概率。

3、利用百度迁徙数据的结果将可以得到地级市之间的人流移动的规模指数。将这些指数与各市之间对应的OD的权重相乘，计算出每个OD对所对应的人流移动的规模指数，通过将每个OD之间的规模指数除以所有OD规模指数的总和，可以计算得出每个OD对的分布概率。

4、利用跨省百度迁徙数据对出行OD进行了修正，具体来说，为了模拟省界附近的道路流量及服务区充电情况，在省界附近道路的北部端点两侧创建了收费站出入口，以模拟跨省流量，并根据缓冲区结果确定站点其所属的省份。进一步利用百度迁徙数据的迁徙指数，计算从各省份到广东省的流量与广东省内流量的比值，并根据收费站所属的省，分别计算以这些收费站为O/D的出行路径流量占省内总流量的比例P，最后计算扩样系数，并修正跨省流量。

## OD结果调整

### 天级到小时级调整（韶关小时级数据）

为了更具体和详细地模拟用户的出行行为，并与真实数据进行对比，可以根据不同时段用户选择出行的概率，将天级的OD出行分布概率调整到小时级别。我们依据韶关市的小时级数据，具体的操作步骤如下：

1. 基于韶关市的小时级数据，计算多天数据的均值，并计算每小时的MR热力值占单日内的热力值的比例。利用得到一天内各小时内用户选择出行的分布概率。
2. 利用得到的单日内每小时的概率分布数据，将其与OD推断得到的单日的分布概率进行相乘，得到每天内每小时OD分布的概率。

### 出行距离调整

由于电动车和燃油车在最大出行距离等方面存在显著差异，特别是在长途出行时，燃油车的比例可能更高。因此，不同出行距离可能会对新能源汽车在总出行车辆中的占比产生影响，应该依据不同的出行距离对新能源汽车的OD分布概率进行修正。当前计划按照每50公里为一个区间，将出行距离划分成多个区段，统计每个区段内新能源汽车占总车辆数的比例，以生成一份出行距离与新能源汽车占比之间的对应数值表。（具体数据尚未确定，之后会进行调整）

在此基础上，可以根据每个OD对之间的最短路径来确定其对应的出行距离，并将其对应到相应的出行距离范围。即将小时级的新能源汽车的OD概率分布，乘以根据出行距离确定的新能源汽车占比，以获得通过出行距离修正后的新能源汽车OD分布概率。

### 新能源车辆占比修正

由于基于MR和百度迁徙数据推测出的车流量及OD实际上是全部类型车的车流量，而新能源汽车可能只占其中的一部分。同时不同城市的新能源车的保有量及其在所有车辆中的占比不同，因此根据不同城市的新能源车保有量的占比对出行分布概率进行了修正。

# 车辆时空轨迹与行为仿真【刘思远】

## ABM仿真技术介绍

Agent-Based Modeling (ABM) 是一种用于模拟和研究复杂系统的计算模型技术。在 ABM 中，系统中的个体被建模为独立的代理(agent)，每个代理有自己的状态、行为规则和与其他代理的交互方式。通过模拟代理之间的相互作用和行为规则，ABM 可以用于研究整个系统的宏观行为。

ABM仿真模型中的关键概念包括代理、状态、行为规则、环境和相互作用。代理是系统中的个体，具有自己的状态和行为规则。状态表示代理在特定时间点的属性或特征，而行为规则定义了代理如何根据当前状态和环境信息做出决策或执行动作。环境是代理交互和行为的背景，可以包括其他代理、物理空间和资源分布等。代理之间通过相互作用影响彼此的状态和行为。通过考虑这些概念，ABM仿真模型可以模拟和分析复杂系统的行为和动态变化，为各个领域的研究提供帮助，例如交通规划、城市规划和生态系统研究等。

ABM仿真过程通常包括初始化、更新代理状态、代理交互和时间推进等步骤。首先，初始化阶段设置仿真环境和代理的初始状态。然后，根据代理的当前状态和环境信息，应用行为规则来更新代理的状态。如果代理之间可以相互作用，执行代理之间的交互过程。接下来，将仿真时间推进到下一个时间步，更新环境和代理的状态。重复执行更新代理状态、代理交互和时间推进的步骤，直到达到预定的仿真结束条件。通过这些步骤的循环迭代，ABM仿真模型可以模拟系统的动态演化和行为变化。

ABM 可以在各种领域中**应用**，例如交通流模拟、城市规划、经济学、生态学等。在车辆时空轨迹和行为仿真中，ABM 可以用于模拟和研究车辆的行驶轨迹、交通流量、交通拥堵、驾驶行为等方面的问题。通过构建每辆车作为一个代理，定义其状态（如位置、速度、目的地等）、行为规则（如遵循交通规则、路径选择策略等）和与其他车辆的交互方式（如避让、协同行驶等），可以模拟并观察整个道路网络中车辆的运行情况和交通流动态。

ABM 的**优势**在于它能够考虑个体之间的异质性、非线性交互和新兴的群体行为。它可以帮助理解和预测复杂系统的整体行为，并为政策制定和规划提供支持和决策依据。然而，ABM 也面临一些挑战，包括数据收集和验证、计算复杂性、参数设置和灵敏性分析等。在进行 ABM 仿真时，需要合理选择和校准模型参数，并进行敏感性分析以评估模型的稳定性和可靠性。总的来说，ABM 是一种强大的模拟技术，可以用于研究车辆时空轨迹和行为，并为交通规划、交通管理和智能交通系统等领域提供决策支持。它能够考虑到个体的异质性和交互作用，从而更好地模拟真实世界中的复杂系统行为。

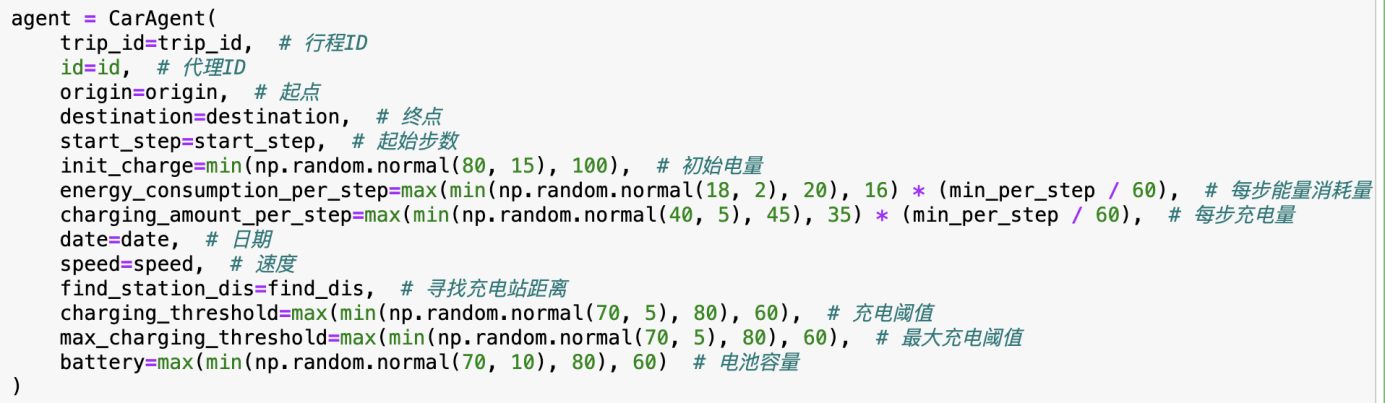
## ABM仿真代理模型

### 车辆代理

当车辆进行**行驶**时，每个车辆代理根据给定的车速和耗电参数，在事先提取好的路网上进行移动。车辆代理记录并跟踪自身的当前位置和电量等信息，以便进行有效的管理和监控。行驶行为的实现涉及几个关键步骤。首先，车辆代理根据预先定义的路网拓扑结构，在每个步骤中选择下一个节点进行移动，从而保证行驶路径的合理性。其次，根据给定的车速参数，车辆代理能够计算出每个步骤中车辆移动的距离和消耗的电量，从而为行驶过程提供准确的能源需求。此外，车辆代理会记录自身的当前位置、电量等信息，并根据行驶状态进行及时的状态更新，以便进行实时监测和调度。

**充电**作为行驶过程中重要的环节之一，也需要得到适当的处理。车辆代理需要根据一定的规则和条件寻找充电站，并进行充电操作。为此，车辆代理会根据当前位置和预设的寻找充电站距离，搜索附近的充电站，并使用一定的策略选择最近的充电站作为充电目标。进一步，车辆代理会根据当前电量和充电需求，判断是否需要进行充电，并根据预设的充电速度参数计算所需的充电电量和充电时长。通过这样的行驶和充电机制，车辆代理能够在路网上高效地行驶，并及时满足充电需求，从而实现可靠的运行和管理。

以下是车辆初始化的示例：



### 充电站代理

充电站代理负责为车辆提供充电服务，并管理车辆的充电需求和状态。充电服务是充电站代理的核心职责之一。充电站代理根据车辆的充电需求和充电速度参数，为车辆提供相应的充电能力。通过考虑充电桩的功率和车辆的充电需求，充电站代理能够计算出充电时长并进行充电操作。在某些情况下，充电站代理可能面临车辆排队等待的情况。当充电站的充电能力有限时，充电站代理需要管理车辆的排队等待状态。为此，在添加车辆到队列时，充电站代理会检查车辆是否已经在等待队列或正在充电队列中，以避免重复添加。若当前充电队列有空余位置，充电站代理会将等待队列中的车辆移到充电队列中，并将其充电状态标记为非行驶状态。通过这样的充电站代理操作，车辆能够有序地进行充电，并且充电站能够合理地管理充电排队等待的情况，以提供高效和可靠的充电服务。

### 代理设计

该模型包含两个代理：汽车代理和充电站代理。汽车代理具有唯一的ID、起点和终点等自身属性，同时记录行驶状态。充电站代理负责维护正在充电和等待充电的汽车队列。模型使用step作为时间单位，表示模拟的每个时间步，通过更新代理状态、执行行动和决策来模拟充电站的运行和汽车的行驶过程。这样的模型可用于充电站的仿真模拟和性能评估，为优化电动车充电系统提供帮助。

CarAgent

1. `run(self)': 当线程开始运行时，该函数被执行。它表示车辆代理的主要逻辑。如果当前节点不是目的节点，车辆代理会更新状态，移动到下一个节点，并可能在附近的充电站充电。如果当前节点是目的节点，车辆代理会被标记为待删除。

2. `charging\_per\_step(self, station\_agent)`: 根据每一步的充电策略，该函数用于在每一步中给车辆充电。如果车辆的电量低于阈值并且处于行驶状态，它会增加电量，并相应地更新充电站的充电需求。

3. `append\_car\_charging\_df(self, station\_id)`: 该函数用于将车辆的充电记录追加到车辆的充电数据帧中。它记录了车辆在特定充电站充电的相关信息。

4. `create\_car\_dataframe(self)`: 该函数用于创建当前车辆状态的数据帧。它收集了车辆的行驶信息，包括当前节点、电量、行驶距离等。

5. `find\_nearby\_charging\_stations(self)`: 返回满足距离要求的当前边附近的充电站。它根据给定的条件筛选出符合要求的充电站，并返回相应的数据。

6. `charge\_at\_nearest\_station(self, nearby\_charging\_stations)`: 在附近充电站中找到距离最近的充电站，并进行充电操作。该函数更新车辆的当前节点，计算需要充电的电量，并调用充电站的相关方法进行充电。

下面是对`CarAgent`类和`StationAgent`类中各个函数的说明：

7. `check\_charging\_station(self, mode)`: 该函数用于判断是否需要充电。如果车辆处于行驶状态且当前节点不是目的节点，它会根据给定的模式检查附近是否有可用的充电站，并决定是否进行充电。

8. `move\_to\_next\_node(self)`: 该函数用于将车辆移动到下一个节点。如果车辆处于行驶状态且当前节点不是目的节点，它会更新车辆的位置、电量等信息，并根据需要决定是否进行充电。

9. `get\_charging\_probability(self)`: 根据车辆的电量情况，计算车辆在当前状态下进行充电的概率。该函数根据车辆的电量调整基础概率，并返回调整后的充电概率。

StationAgent

1. `\_\_init\_\_(self, charge\_node\_id, max\_capacity)`: 初始化充电站代理对象。设置充电站的节点ID、最大容量、剩余容量、等待队列、当前队列、充电需求、车辆数量和数据帧等属性。

2. `if\_not\_exit\_in\_queues(self, car\_agent)`: 检查车辆代理是否不存在于当前队列和等待队列中。返回布尔值，表示车辆是否不存在于队列中。

3. `append\_car\_charging\_df(self, car)`: 将车辆的充电记录追加到充电站的充电数据帧中。创建一个包含车辆充电信息的字典，并将其转换为数据帧格式返回。

4. `append\_to\_stationDf(self)`: 将当前充电站的信息追加到充电站数据帧中。创建一个包含充电站相关信息的数据帧，并将其与之前的数据帧合并。

5. `has\_capacity(self)`: 检查当前充电站是否有剩余容量。返回布尔值，表示当前充电站是否还有空余位置。

6. `charge\_cars(self, car\_agent)`: 对当前队列中的车辆进行充电。如果车辆在当前队列中，调用车辆代理的`charging\_per\_step`方法进行充电。

7. `add\_to\_queue(self, car\_agent)`: 将车辆代理添加到充电站的队列中。根据车辆代理是否存在于队列中和充电站是否有空余位置，将车辆添加到当前队列或等待队列，并更新相应的数据帧和车辆数量。

8. `move\_cars\_from\_waiting\_to\_current(self)`: 将等待队列中的车辆移动到当前队列。当当前队列有空余位置且等待队列不为空时，从等待队列中取出车辆并添加到当前队列中，并更新相应的数据帧和车辆数量。

9. `station\_run(self, car\_agent)`: 充电站的执行逻辑封装。该函数接受一个car\_agent参数，表示要处理的车辆代理。首先，将车辆代理的状态设置为非行驶状态（state=0）。然后，调用add\_to\_queue方法将车辆代理添加到充电站的队列中。接下来，调用charge\_cars方法对当前队列中的车辆进行充电操作。如果充电站有空余位置，会从等待队列中将车辆移动到当前队列中，直到当前队列达到最大容量或等待队列为空。

下面是对`Stations`类中各个函数的说明

## 仿真模型参数与场景设置

### 仿真模型参数：

### Screenshot 2023-10-30 at 16.19.25

### 场景设置：

在建立充电站仿真模型时，还需要考虑以下几个方面。对于高速场景下的路网结构比较单一的情况，可以选择使用最短路径算法来确定车辆的行驶路径。这意味着车辆在仿真中选择的路径将是最短的路线。充电逻辑的决策可以基于电量作为考量因素。例如，可以使用以电量为均值的正态分布作为充电结束的阈值。当车辆的电量达到或超过阈值时，仿真模型可以将其安排去充电站进行充电。在确定车辆数量和分布时，可以采用一天一万辆车左右的数量。为了更真实地模拟车辆的分布情况，可以根据求解得到的广东省内起点和终点的车流量数据，计算概率分布，并进行采样来确定车辆的起点和目的地。充电站的设置可以基于真实的充电站情况进行。这包括充电车位数和充电站的位置。可以根据实际充电站的信息，确定充电站的数量和位置分布，以便在仿真中准确地模拟充电站的布局和可用资源。通过综合考虑路网结构、充电逻辑、车辆数量和分布以及充电站的设置，可以建立一个更真实和可靠的充电站仿真模型，以便进行相关研究和分析。

### 模型优势：

根据高速公路场景的特点，与城市场景相比，充电决策更为简化。在高速场景中，充电决策主要是为了确保车辆能够到达目的地，而不受工作或其他需求的影响。因此，模型将充电决策仅与当前电量以及最近的充电站之间的距离相关联，从而简化了逻辑。

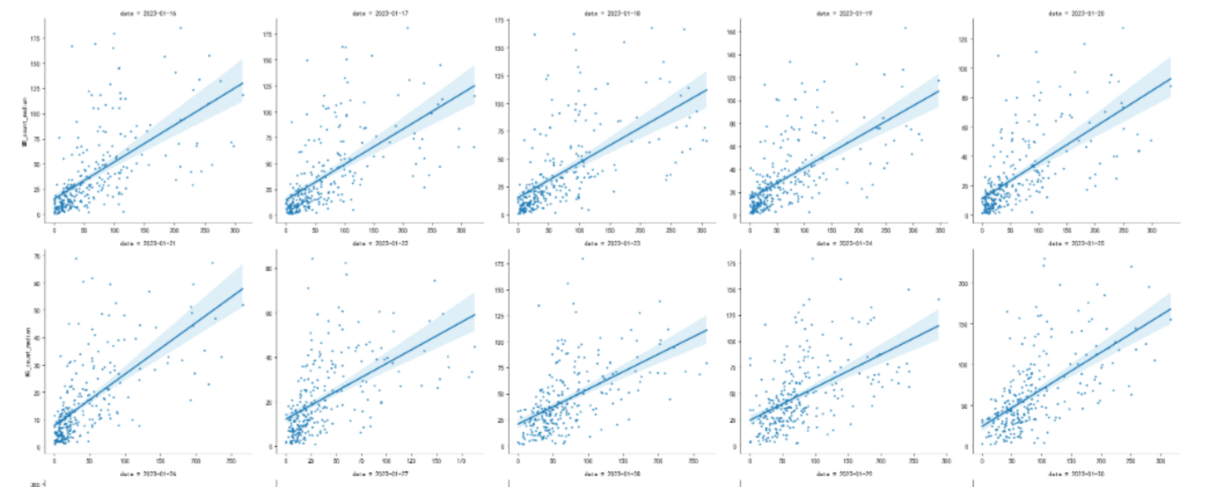
这种简化的逻辑有助于提高模型的运行效率，并使其更加适应高速行驶的要求。车辆需要快速做出充电决策以保证行程的顺利进行，而复杂的决策逻辑可能导致计算和操作的延迟，从而影响到车辆的性能和可靠性。

简化后的ABM仿真模型更加注重于刻画每个时间步的车辆和充电站信息，并从原始数据中提取更多有用的信息。例如，可以获取**上下游路段**的车辆信息以及**高位、低位**充电车辆的比例等。这些数据为充电站选址提供了数据支持。通过考虑上下游路段的车辆信息，能够更准确地了解充电站周围的交通状况和需求情况。这为确定充电站的最佳位置提供了重要的参考依据，以满足车辆充电需求并确保充电站的可达性和利用率；通过分析高、低位充电车辆占比等信息。这些数据能够揭示充电需求的特点和趋势，例如高位充电需求更高的区域或时间段。基于这些信息，能够更精确地确定充电站的数量、位置和配置，以最大程度地满足高速场景下的充电需求，并优化充电站的利用率和效益。

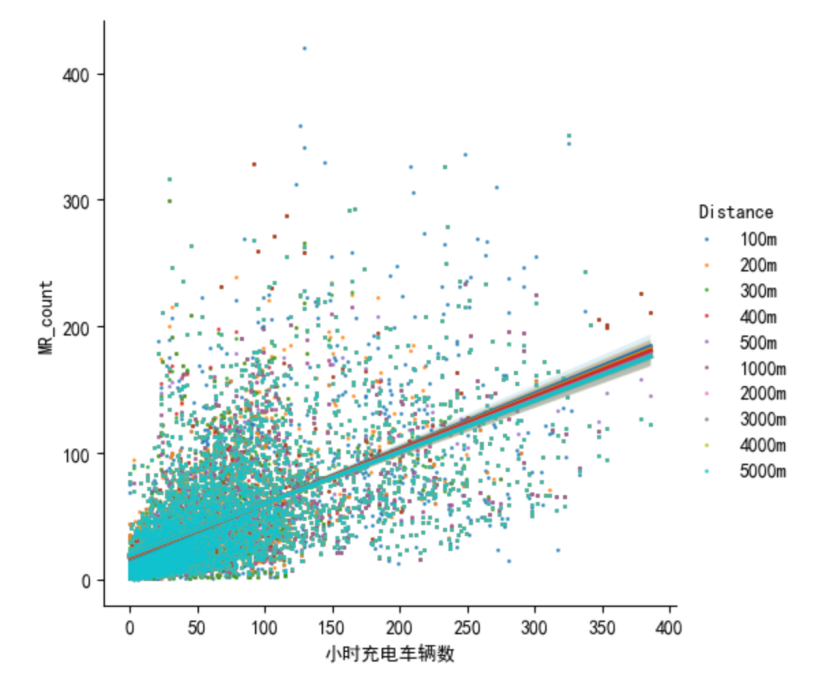
# 服务区充电需求预测评价与场站配置建议【李家兴】

## 上游路段MR与充电车辆数比较结果

将站点上游路段MR与站点小时级充电车辆数进行对比分析，站点小时级充电车辆数和站点上游路段的MR显示出一定的相关性。通过分析不同日期的数据，均呈现出一定的相关性，相关系数达到0.6。小时级内较多的MR数据可能意味着较多的车流量，车流量较大时，充电需求也将较大，MR能够一定程度反映高速道路运行状况，基于MR数据推测出行OD具有可行性。同时表明，考虑充电站选址及扩容时，可考虑能反映车流量、上游路段MR等的指标。



分析不同上游路段范围内（100米至5000米）的MR与站点小时级充电订单的相关性，发现不同距离范围对小时级充电订单的相关性影响明显。总体而言，相关系数保持在0.58至0.61之间。200米范围内的上游路段与小时级充电订单的相关性最为显著，相关系数达到0.61。表明距离较近的路段的MR对充电需求的影响更加显著。随着距离范围的增加，相关系数逐渐减少，意味着随着距离的增加，上游路段MR对充电订单的影响逐渐减弱。在使用MR数据时，可考虑上游路段200米范围的MR作为分析指标。

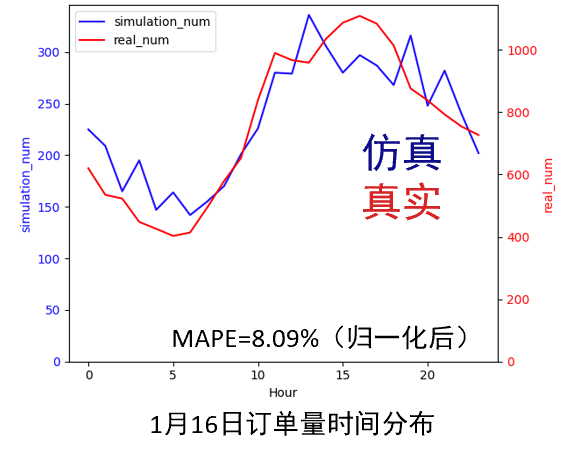
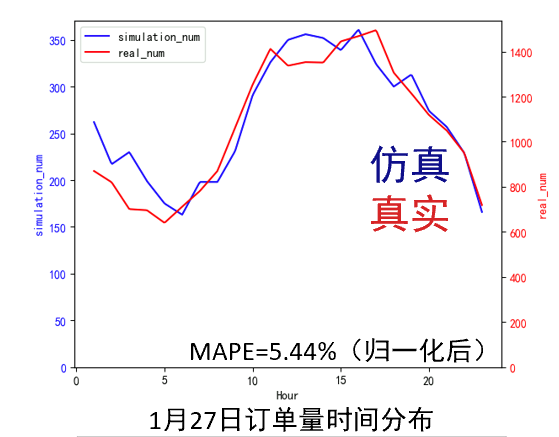


## 仿真的结果比较分析

仿真结果比较将选取订单量这一指标进行对比。

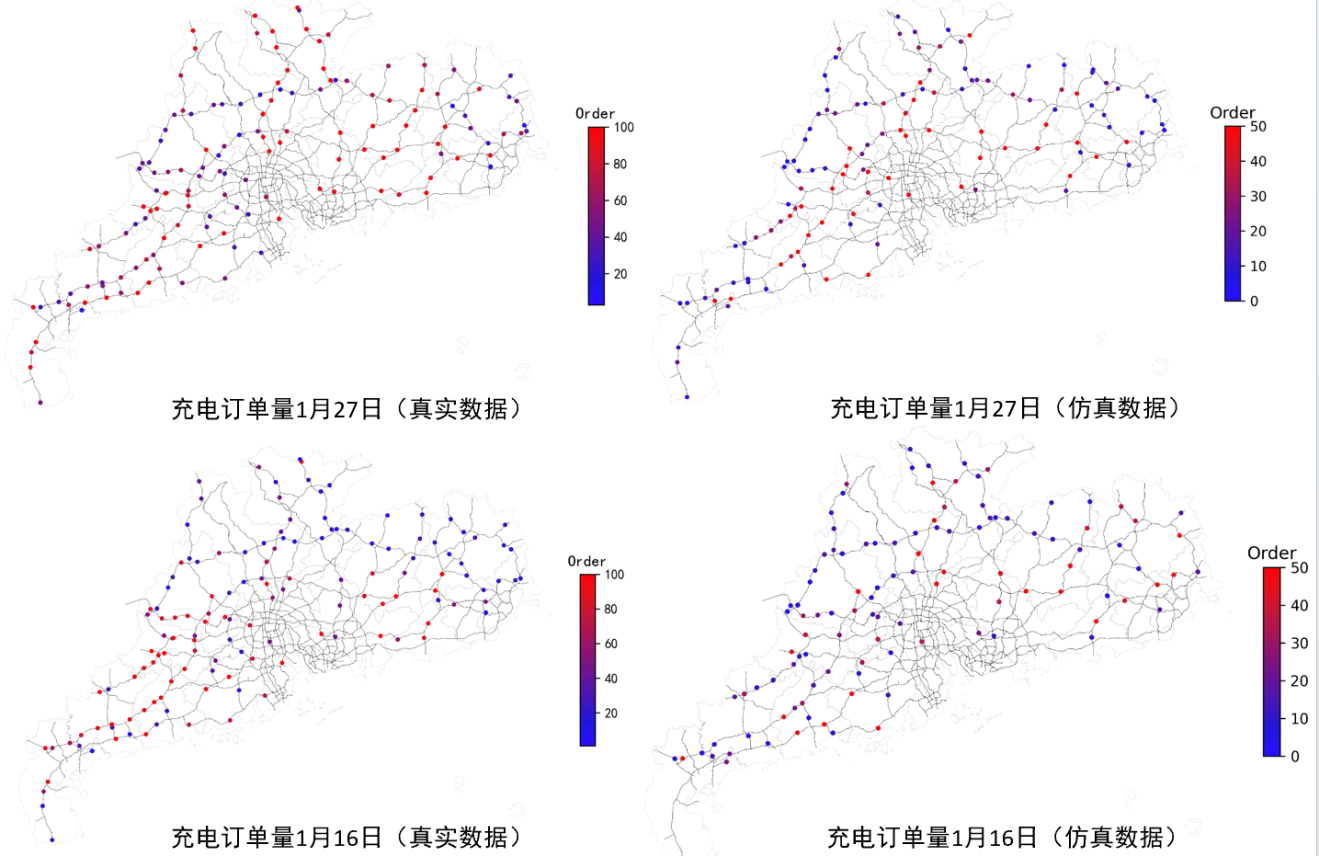
**（1）小时级订单量**

选取一天对比仿真结果和真实结果的小时级总订单量曲线时，发现二者呈现出较高的一致性。尤其在夜间，充电订单数量相对较少，而在12点至15点之间出现了充电高峰。表明仿真结果与现实场景拟合较好。仿真结果的夜间的充电需求较低，但在白天特别是中午到下午，出现了显著的充电高峰期，与真实结果十分相似。



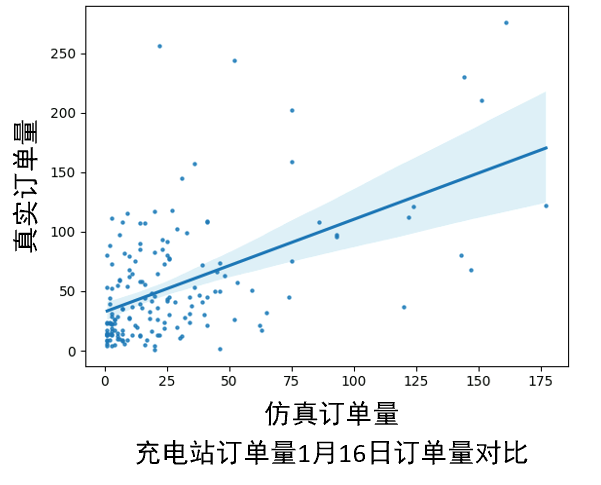
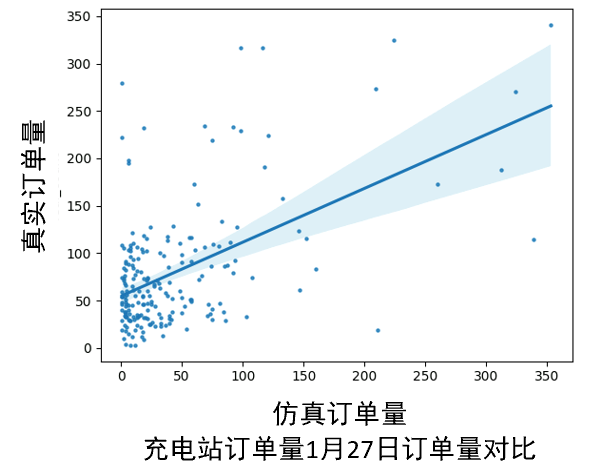
**（2）站点总订单量空间分布**

对比一天内仿真结果和真实结果的站点级总订单量空间分布时，发现总体的空间分布趋势较为一致。在粤西地区，例如云浮、肇庆和阳江等地的服务区，充电订单数量较为集中。相对而言，在粤东地区，河源至潮汕方向的服务区充电订单数量较多。而大湾区外围区域，充电订单数量相对较少。



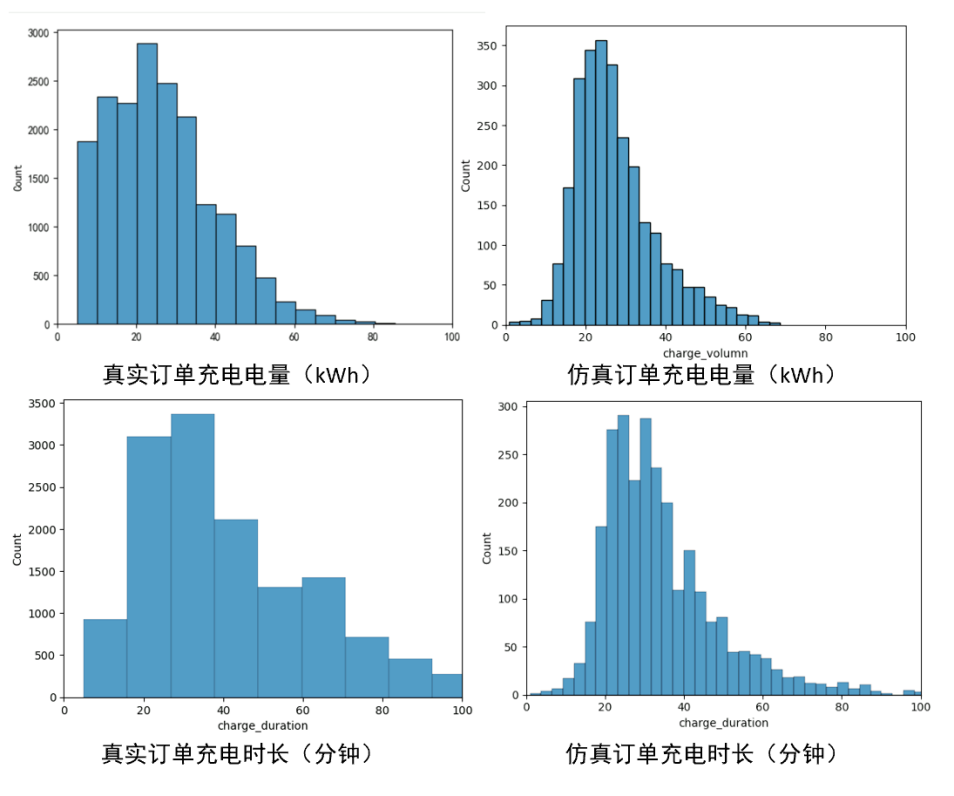
**（3）站点总订单量拟合程度**

对比站点级的总订单量仿真结果与真实结果，绘制拟合曲线，比较仿真结果与真实结果。部分站点的真实订单数量相对较小，仿真结果的订单数也较小，说明模型对这些站点的仿真较为准确。然而，也可能存在一些站点，其仿真订单数量低于真实数据。可能由于仿真所使用的车辆数与真实情况存在较大的量级差异，导致仿真的这些站点的订单量较少，这些站点在散点图的上方。



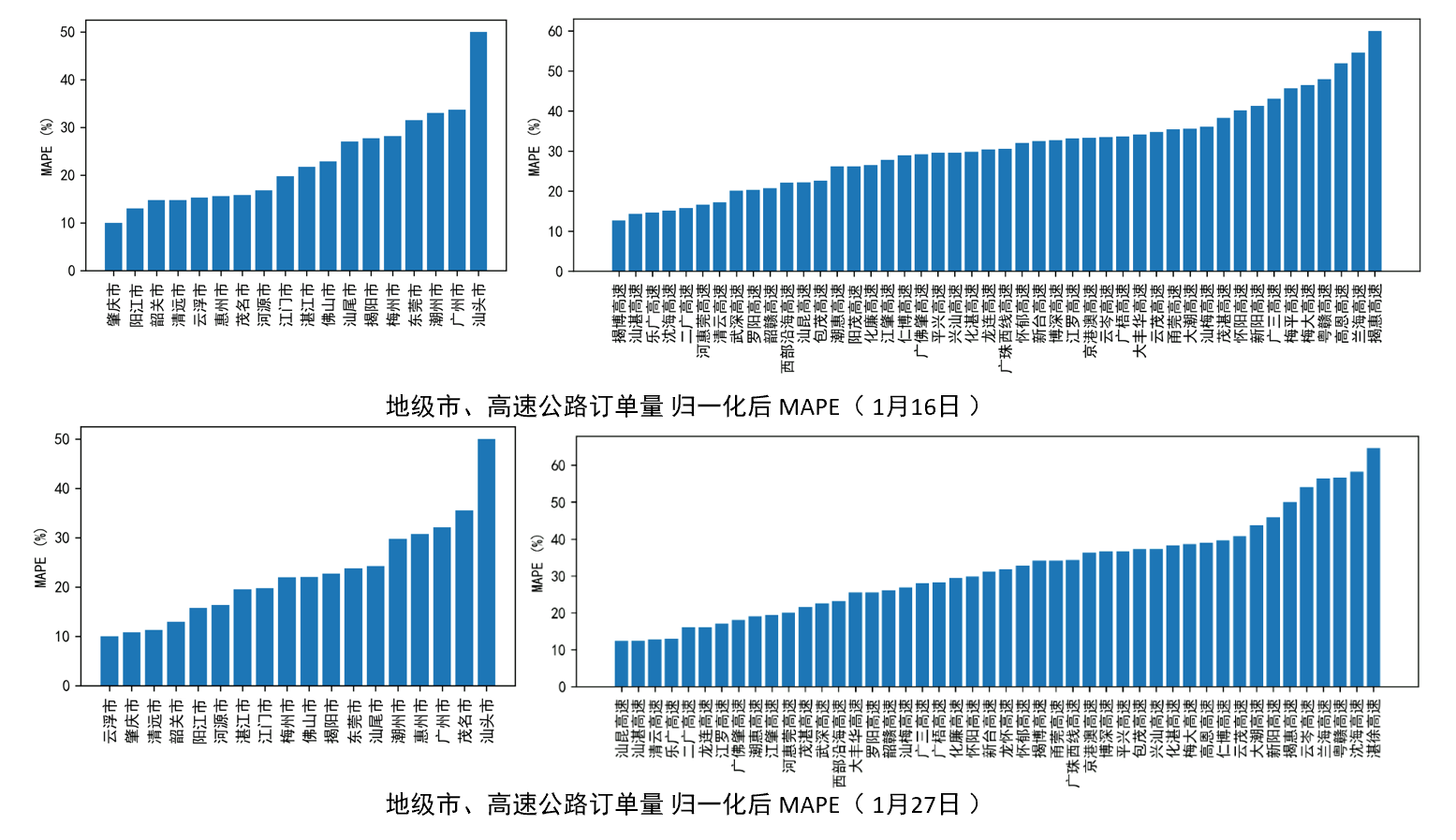
**（4）订单充电量分布**

对比仿真与真实数据的电量分布和充电时长呈现出总体相似的趋势，其中订单电量均值集中在20-30 kWh之间，充电时长主要分布在20-30分钟的区间。这种相似性表明仿真模型在模拟电动车充电过程中的整体趋势方面表现良好。



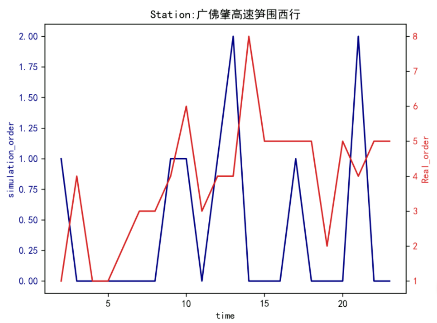
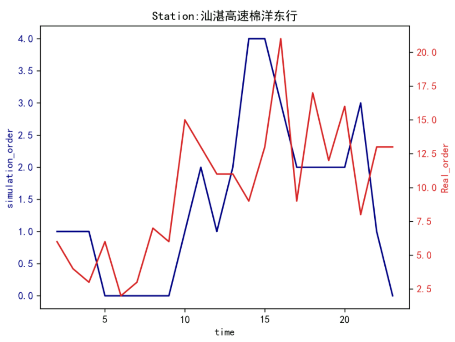
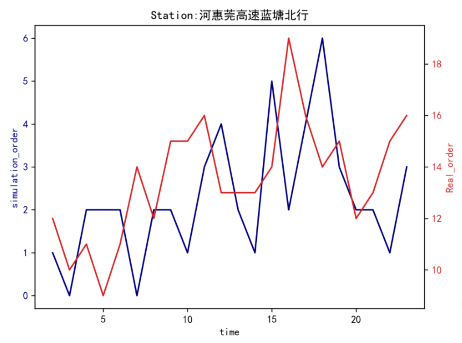
**（5）城市、高速公路级仿真精度**

分析城市级、高速公路级仿真精度。部分城市如肇庆、云浮、清远等城市精度较高、汕头、茂名等城市精度较低。分析道路级别的仿真进度，汕湛高速、乐广高速的精度较高，揭惠高速、兰海高速的精度较低。



**（6）部分站点小时级订单量**

选取部分站点小时级仿真结果与真实订单进行对比，仿真结果的小时级订单总体趋势与真实订单较为相较，夜间订单数量较少，高峰小时的订单量趋势也较为一致。在特定的时间段内或特定的站点上，实际结果和仿真结果可能会有一些细微的差异，但总体上能够看出仿真结果与实际情况之间的相似性。



## 选址参考指标

**1、站点整体特征：**

（1）充电站订单量： 站点一天内小时级的充电订单量和一天的订单总量。这个指标揭示了站点的充电需求，利用仿真结果数据和真实订单数据进行分析，可以识别时间上的周期性的需求变化。选址时，高订单量的站点可能需要更多的充电桩来满足需求。

（2）充电需求： 站点一天的总充电需求（kwh），为该站点一天订单的充电量总和。该指标将评估站点的整体充电需求量，有助于确定站点的承载能力。对于选址，需要考虑站点是否有足够的能力满足现有的和潜在的充电需求。

（3）充电桩需求缺口：仿真充电需求与真实充电需求间差值。衡量实际充电需求与当前充电桩数量之间的差距，一方面可以确定建设充电桩的紧迫性，另一方面可以指导充电桩布设数量。

（4）充电桩利用率： 充电桩一天内的使用率，为充电桩使用时间占一天24小时时间的比值。分析充电桩的实际利用情况有助于了解是否存在充电桩利用不足或过度的情况。确定是否需要重新调整资源配置，以更好地满足充电需求。

**2、车辆个体特征：**

（1）进入充电站时电量分布： 分析车辆进入充电站时的电量（kwh）分布。反映了不同电量需求的车辆数量了解车辆充电需求的分布特征，以确定适当的充电桩类型。

（2）服务区充电时长： 车辆在服务区充电的平均时长。评估车辆在充电站的停留时间。

（3）服务区排队时长：车辆在充电站的平均排队时长。可以确定充电桩是否需要进行扩容，并对充电桩布设前后排队时长进行分析以评估选址方案。

（4）订单的充电电量分布：站点级的充电订单充电量分布。

…