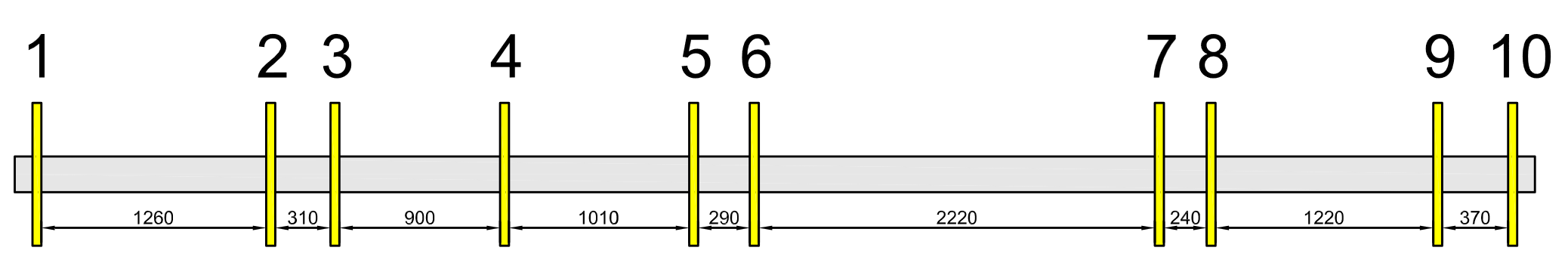
**公路交通流量预测数据集说明**

**一、检测器分布**

道路的检测器编号和分布如下所示，共十个检测器，其中检测器间距离标注如下，单位为米（m），精度为10米。道路为单向四车道，运行方向为从1→10。



**二、数据集**

数据集包含Train-5min，Test-5min，Train-30s，Test-30s共四个数据包，每个数据集包含1-10共十个csv文件，分别对应相应编号的检测器数据。

其中Train-5min，Test-5min数据以5分钟为单位进行集计。csv含五个基础字段：时间（Time）、流量（Flow，veh/h）、速度（Speed，km/h）、占有率（Occupancy，%）、星期（Week）。训练集（train）数据包含90天全天的检测器数据，测试集（test）包含7天5:00-7:55、9:30-12:25、14:00-16:55的检测器数据，两者数据在时间上连续。该数据集请注意以下几点：

（1）数据时间已经统一处理到开始为00:00，实际数据起始日期不予提供。

（2）数据存在1%左右的随机缺失，缺失部分为nan，请自行填补。同时存在一定数量的错误数据，请自行甄别。

（3）数据相比原始数据，已置入一定值的高斯白噪声。

Train-30s，Test-30s数据集为Train-5min，Test-5min原始数据，检测时间间隔为30s，其中流量单位为Veh，表示30s内通过检测器的车辆数；速度单位为mph。该数据仅作为参考使用。

**三、预测任务**

使用测试集数据，对测试集各天8:00-8:55、12:30-13:25、17:00-17:55的所有检测器的流量和速度进行预测。

最后提交任务范例见文件“提交范例.csv”。最后提交文件为csv格式，时间与星期已给出，预测结果按照检测器1到检测器10分布，每个检测器需要预测流量和速度。预测流量为四车道相加值，保留整数，速度为每车道流量加权平均，保留两位小数。流量和速度计算公式如下所示：

其中为单一车道；为单一车道流量；为单一车道速度。

**四、评价指标**

测试误差评价指标及计算方法如下所示：

其中为选定的预测变量，即流量和速度；为需要预测的时间段，即测试集中7天共21个时间段；为选用的10个检测器。Aver、MAE、RMSE的计算公式如下所示：

其中为数据数量，为真实值，为预测值。