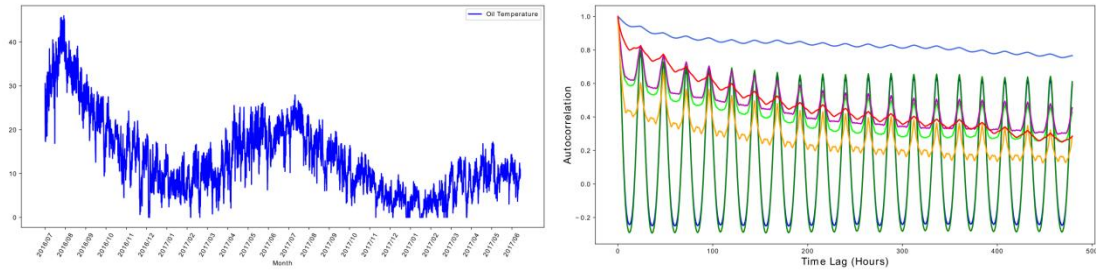


### 数据介绍：

本测试中所使用的数据集为电力变压器数据集 ETTh1.csv，时间跨度为 2016 年 7 月到 2018 年 7 月共 2 年的数据，数据的具体含义如下：

Field	date	HUFL	HULL	MUFL	MULL	LUFL	LULL	OT
Description	The recorded date	High UseFul Load	High UseLess Load	Middle UseFul Load	Middle UseLess Load	Low UseFul Load	Low UseLess Load	Oil Temperature (target)

其中主要变量是变压器油温，油温的总体概览与各变量的概览如下图所示：



油温特征在数据集中总览；全部变量的自回归图形展示

### 时间序列回归任务：

本任务要求对该多维时间序列进行回归预测，即利用历史信息对未来进行预测，具体的任务可以拆解成如下几个组成部分

#### 1. 数据处理

标准化是常用的数据处理方式，本任务要求使用 z-score 标准化，利用 2016.7-2017.6 的数据统计量，对整个数据集进行标准化，标准化后的数据可表示为  $\{\mathbf{x}_t | \mathbf{x}_t \in \mathbb{R}^7\}_t$ ，为 7 维的变量。（注意模型评估也在标准化后的数据上进行）

#### 2. 构建预测模型

根据历史回看窗口长度  $T$  的数据  $\mathbf{x}_{t-T < t' \leq t} \in \mathbb{R}^{7T}$ ，预测未来时间窗口长度为  $H$  的数据  $\mathbf{x}_{t < t' \leq t+H} \in \mathbb{R}^{7H}$ 。即需要构建模型  $f_\theta$ ，对未来数据进行预测。

$$\hat{\mathbf{x}}_{t < t' \leq t+H} = f_\theta(\mathbf{x}_{t-T < t' \leq t})$$

其中模型  $f_\theta$ ，可以使用任意模型，历史回看窗口长度  $T = 336$ ，预测窗口长度为  $H = 24$ 。即在每个时间点上，模型回看 336 小时（14 天）的数据预测未来 24 小时的数据。

#### 3. 数据划分与滚动

本任务的预测目标是预测 2017.07.01 00:00:00（包含）之后的数据，并采用每月滚动更新训练的形式：即要求在 2016.7-2017.6 上训练模型，并且预测 2017.7 的数据，后加入 2017.7 的数据进入训练集合再训练模型，在 2017.8 上

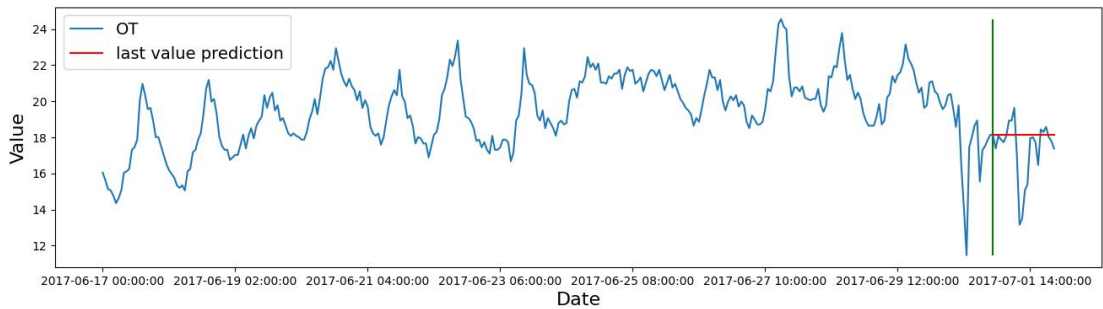
进行预测，依次类推。

4. 效果评估

模型评价以 MSE 和 MAE 误差作为评价标准，要求得到每次滚动模型评估的效果。

结果呈现，要求画出模型预测结果表现（如下图表示，绿色线左侧表示回看窗口，右侧表示预测窗口，蓝色曲线表示真实值，红色曲线表示预测值），下图展示了“利用上一个值作为预测结果”的简单预测方法，该方法可以作为简单的基线模型，用于对比所实现方法的性能。

预测结果展示图，对于每个变量展示几张预测结果即可。



5. 实验报告

将以上实验思路和结果形成实验报告，要求说明算法思路和主要指标的结果及模型评估部分相关图表。