

# 基于 Transformer-CNN 模型的市区电力消耗预模型

## 1 github 链接

<https://github.com/kkkl1234/python-.git>

## 2 问题概况

### 2.1 问题背景

电力消耗预测是区域能源调度、资源优化配置的重要支撑，其预测精度与波动率捕捉能力直接影响电力系统的稳定运行与效率提升。摩洛哥北部城市丹吉尔的电力消耗受季节变化、环境因素及区域功能差异的综合影响，呈现出显著的阶段性波动特征；同时，该市三个配电网区域的能耗数据存在潜在的关联规律，这对时序预测模型的综合适配能力提出了更高要求。传统单一建模思路难以同时兼顾时序趋势捕捉、外部影响因素融合及跨区域关联挖掘，无法充分适配丹吉尔电力消耗的复杂变化特征，亟需针对性的模型优化与方法创新。

### 2.2 问题提出

经典 ARIMA 模型虽能有效拟合平稳时序数据的内在规律，但面对电力消耗数据的非平稳性、波动突变特性，以及温度、湿度、漫射辐射等外部环境因素的干扰时，模型泛化能力不足；此外，现有研究中单一指标或单一区域的建模方式，既忽略了关键环境特征对能耗的驱动作用，也未挖掘区域间能耗的协同关系，导致预测精度受限。同时，电力消耗波动率（尤其是极端值波动）缺乏有效的量化分析方法，难以支撑电力系统风险预警与精细化调度的实际需求。在此背景下，亟需构建适配复杂影响因素的电力消耗预测模型，解决现有方法的短板问题。

### 2.3 问题描述

本研究以丹吉尔 2017 年 1 月 1 日至 12 月 30 日的电力消耗时序数据（涵盖 3 个区域能耗指标）及 5 项环境监测特征（温度、湿度、风速、两类漫射辐射）为研究基础，聚焦以下核心问题展开分析：

- 如何构建兼顾时序趋势与波动特征的预测模型，突破传统模型对非平稳、突变性时序数据的适配瓶颈，提升电力消耗的样本外预测精度。
- 如何建立电力消耗波动率的量化分析方法，为极端能耗事件预警提供技术支撑。

研究将通过对比 ARIMA 模型与 Transformer-CNN 混合模型的预测性能，验证单指标、多特征、多区域联合建模的效果差异，最终为丹吉尔区域电力消耗精准预测与调度决策提供科学依据。

## 3 数据概况

### 3.1 数据来源

数据集来自 UCI 公开数据集官网，包含摩洛哥北部城市丹吉尔自 2017-01-01 至 2017-12-30 日每天以十分钟为节点，三个不同配电网的电力消耗情况以及观经检测特征。包含环境监测特征 5 项，分别为温度、湿度、风速、总漫射辐射、漫射辐射，均为影响建筑能耗的关键外部环境变量。

### 3.2 数据探索性分析

电力消耗指标 3 项：1/2/3 区电力消耗，以及与电力消耗相关的五个环境指标，具体数据以下表为例：

表 1: 区域电力消耗时序数据 (2017-01-01)

DateTime	Zone 1 Power Consumption	Zone 2 Power Consumption	Zone 3 Power Consumption
2017-01-01 00:00:00	34 055.696	16 128.875	20 240.963
2017-01-01 00:10:00	29 814.683	19 375.075	20 131.084
2017-01-01 00:20:00	29 128.101	19 006.686	19 668.433
2017-01-01 00:30:00	28 228.860	18 361.094	18 899.277

表 2: 环境因素时序数据 (2017-01-01)

DateTime	Temperature	Humidity	Wind Speed	general diffuse flows	diffuse flows
2017-01-01 00:00:00	6.559	73.8	0.083	0.051	0.119
2017-01-01 00:10:00	6.414	74.5	0.083	0.070	0.085
2017-01-01 00:20:00	6.313	74.5	0.080	0.062	0.100
2017-01-01 00:30:00	6.121	75.0	0.083	0.091	0.096

由于数据集数据是每十分钟更新，且并无缺失值，因此文章首先将以天级为单位进行数据处理，通过对每天的数据取平均值，作为当天的数据，处理后的数据如下所示。

表 3: 区域电力消耗时序数据（处理后）

DateTime	Zone 1 Power Consumption	Zone 2 Power Consumption	Zone 3 Power Consumption
2017-01-01	28 465.232	17 737.791	17 868.795
2017-01-02	28 869.493	19 557.725	17 820.763
2017-01-03	30 562.447	20 057.269	17 620.803
2017-01-04	30 689.831	20 102.077	17 673.694

表 4: 环境因素时序数据（处理后）

DateTime	Temperature	Humidity	Wind Speed	general diffuse flows	diffuse flows
2017-01-01	9.675	68.519	0.315	121.390	25.993
2017-01-02	12.476	71.456	0.076	120.404	27.227
2017-01-03	12.100	74.981	0.076	120.686	28.574
2017-01-04	10.509	75.459	0.082	122.959	28.827

### 3.3 数据可视化

对三个地区的电力消耗指标进行可视化，结果如图 1 所示：

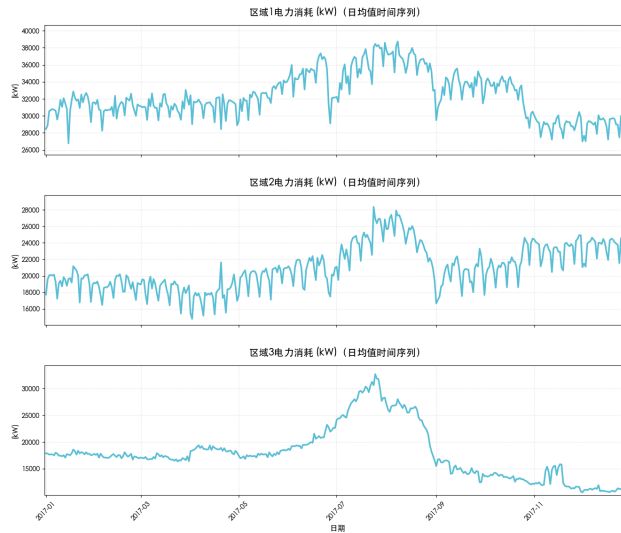


图 1: 各区域电力消耗时间序列图（2017-01-01 至 2017-12-31）

5 个环境特征的时间序列图如图 2 所示：

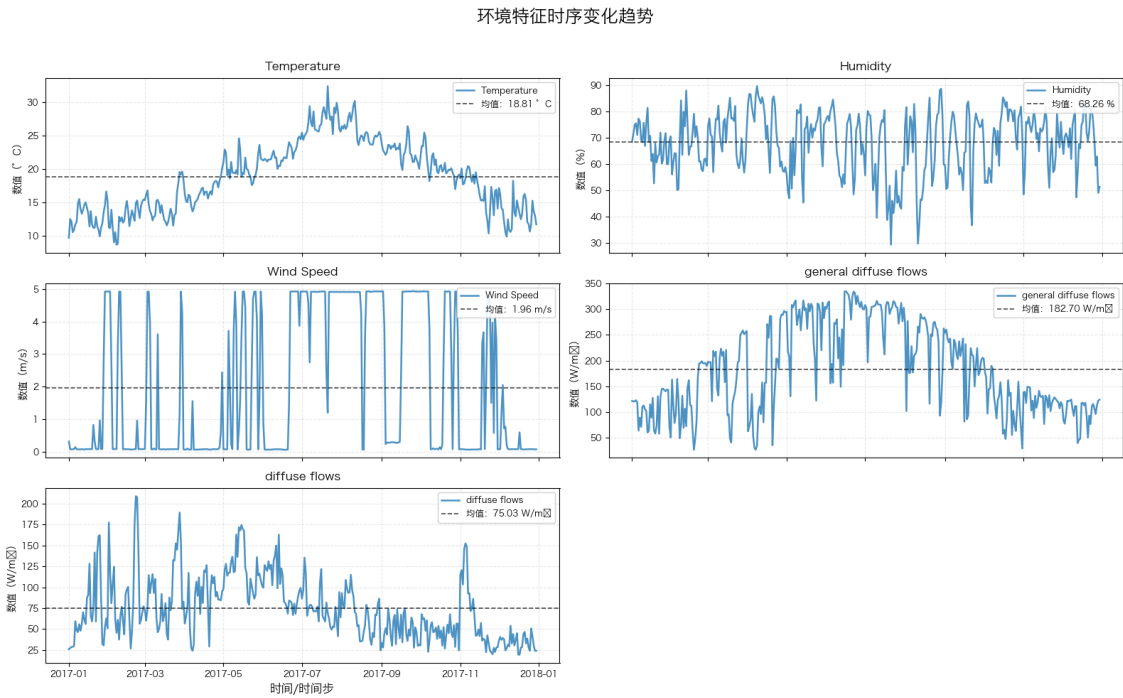


图 2: 温度等 5 个环境变量的时间序列图

由时间序列图可以看出，三个区域的能耗均呈现明显的阶段性波动：从趋势性来看，三个地区 2017 年初至 7 月左右呈上升趋势，7 月后出现大幅下降，后续维持在相对低位波动，推测与季节（如夏季制冷需求）或区域功能调整有关。从周期性来看，三个地区未呈现规律的、重复的周期波动，仅能看到“短期小幅震荡”，但这种震荡无稳定的周期长度，因此周期性不显著。从局部波动性来看，区域 1、2 的波动幅度更大，区域 1 峰值近 4 万 kW，区域 2 峰值约 2.8 万 kW，区域 3 前期波动平缓、后期降幅更显著。

从五个环境特征的时间序列图可以看出：温度整体呈单峰型趋势，先持续上升至峰值后逐步回落，均值 18.81°C；湿度波动平缓且与温度反向变动，均值 68.26 风速无长期趋势，呈无规律脉冲式剧烈波动，均值 1.96m/s；两类漫射流趋势高度同步，均表现为“平稳 - 波动 - 回落”的阶段性变化，均值分别为 182.70W/m<sup>2</sup>、75.03W/m<sup>2</sup>。

## 4 模型建立

### 4.1 ARIMA 模型

#### 4.1.1 模型介绍

自回归积分移动平均模型 (ARIMA) 是时间序列分析中经典的线性预测模型，核心用于捕捉非平稳时间序列的自相关趋势与波动特征，其模型形式记为  $ARIMA(p, d, q)$ ，其中  $p, d, q$  分别代表自回归阶数、差分阶数与移动平均阶数，三参数共同构成模型的核心框架。

1. 自回归部分 ( $AR(p)$ ): 刻画当前值与前  $p$  期历史值的线性关联，表达式为:

$$y'_t = \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \varepsilon_t$$

其中  $y'_t$  为差分后平稳序列， $\phi_i (i = 1, 2, \dots, p)$  为自回归系数， $\varepsilon_t$  为白噪声残差;

2. 差分部分 ( $I(d)$ ): 对非平稳原始序列  $y_t$  进行  $d$  阶差分得到平稳序列  $y'_t$ ， $d = 0$  表示序列本身平稳， $d = 1$  为一阶差分 ( $y'_t = y_t - y_{t-1}$ )，是最常用的差分形式;
3. 移动平均部分 ( $MA(q)$ ): 刻画当前残差与前  $q$  期残差的线性关联，结合  $AR(p)$  的完整模型为:

$$y'_t = \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

其中  $\theta_j (j = 1, 2, \dots, q)$  为移动平均系数。

ARIMA 模型的前提是差分后序列满足“弱平稳性” (均值、方差、自相关系数不随时间变化)，且残差  $\varepsilon_t$  为独立同分布的白噪声 (均值为 0、方差恒定、无自相关)。

#### 4.1.2 模型建立

针对区域 3 的电力消耗时间序列，其呈现出明显的趋势特征。为验证序列平稳性，运用 ADF 检验对原始序列进行检验，结果显示检验统计量对应的  $p$  值为 0.5401 (大于显著性水平  $\alpha=0.05$ )，可判定原始时间序列为非平稳序列。

为消除序列非平稳性对后续建模的影响，对原始序列进行一阶差分变换。随后对所得一阶差分序列再次执行 ADF 检验，结果显示  $p$  值降至 0.0002 (远小于  $\alpha=0.05$ )，可认为差分序列满足平稳性要求，可用于后续时间序列模型的构建与分析。

针对区域 1 经一阶差分处理后的平稳电力消耗序列，构建 ARIMA 模型以实现序列拟合与后续预测。模型阶数选择综合基于两大核心依据：一是 AIC 准则，该准则通过平衡模

型拟合优度与复杂度以筛选最优模型；二是一阶差分序列的自相关函数（ACF）图与偏自相关函数（PACF）图的特征分析如图 3 所示。

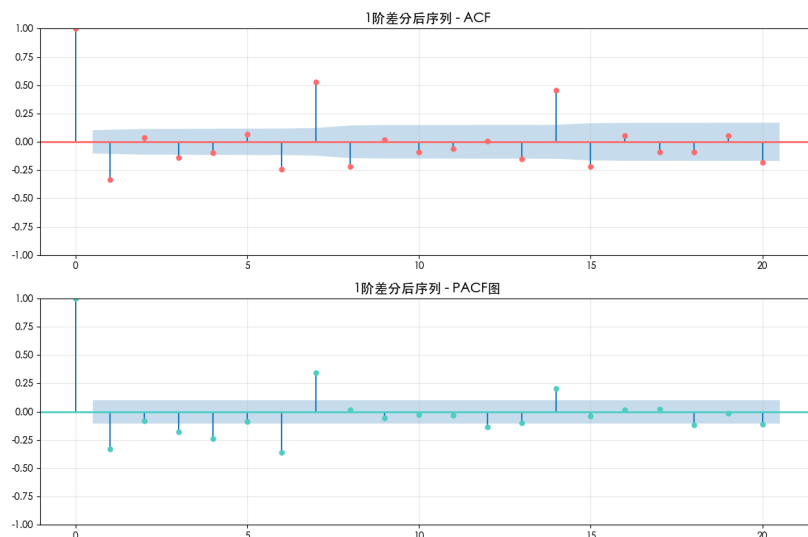


图 3: 区域 3 序列一阶差分 ACF 图与 PACF 图

经多组候选模型的对比验证，ARIMA (2,1,2) 模型对应的 AIC 值最小，PACF 图中，滞后 2 阶以内的偏自相关系数仍未完全稳定落入置信区间，且 ACF 图中，滞后 2 阶以内的自相关系数仍有弱超出置信区间的情况，因此确定该模型为最优拟合模型。为验证模型合理性，对 ARIMA (2,1,2) 模型的预测结果如图 4 所示：

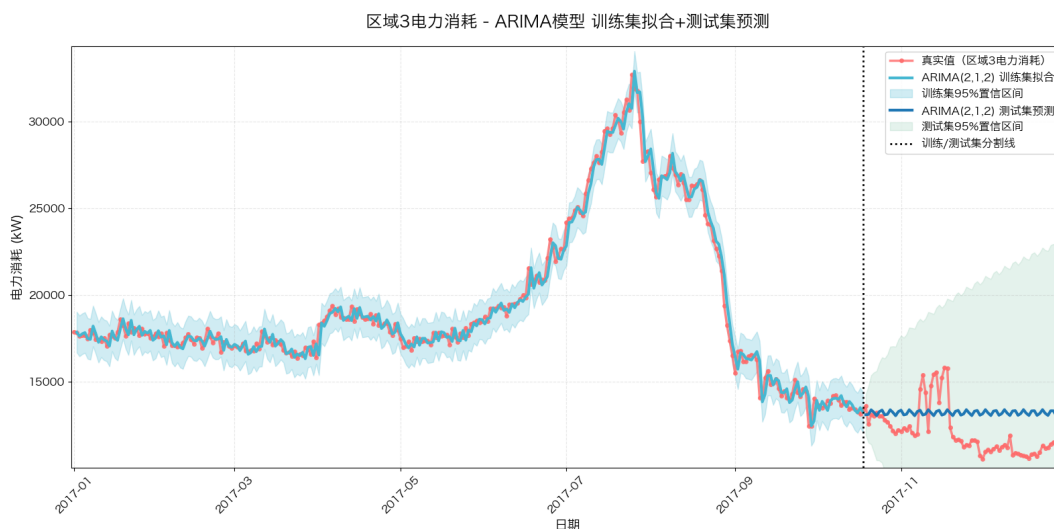


图 4: ARIMA(2,1,2) 模型预测图

模型的参数预测如下表所示：模型的相关预测结果如下表所示：

表 5: ARIMA 模型系数估计结果

系数项	coef	std err	z	P>  z	[0.025	0.975]
ar.L1	-0.4557	0.010	-46.330	0.000	-0.475	-0.436
ar.L2	-0.9998	0.002	-444.451	0.000	-1.004	-0.995
ma.L1	0.4597	0.034	13.564	0.000	0.393	0.526
ma.L2	0.9997	0.062	16.228	0.000	0.879	1.120
sigma2	342900.0	0.000187	1.83e+12	0.000	342900.0	342900.0

表 6: ARIMA 模型电力消耗预测误差评估结果

数据集	有效点数	MAE (kW)	RMSE (kW)	MAPE (%)
训练集 (样本内)	290	402.75	558.50	2.09
测试集 (样本外)	73	1632.39	1782.03	13.89

### 4.1.3 ARIMA 模型预测效果分析

测试集阶段，模型的预测精度出现显著下降。可视化结果显示，测试集区间内真实值序列呈现出训练集未覆盖的剧烈波动特征，而 ARIMA 预测曲线（蓝绿色曲线）趋势相对平缓，未能有效跟踪真实值的动态变化，且测试集 95% 置信区间（浅绿色填充区域）宽度大幅增加；误差指标也印证了这一现象：测试集 MAPE 攀升至 13.89%，MAE 与 RMSE 分别升至 1632.39 kW 和 1782.03 kW，表明模型对未见过的数据泛化能力较弱，难以适配训练集外的波动模式突变。

#### 4.1.3.1 整体结论

综合来看，本研究构建的 ARIMA 模型呈现“样本内拟合优、样本外泛化弱”的典型特征：其优势在于能够有效挖掘历史数据中的时间序列规律，适用于趋势稳定、波动模式单一的短期电力消耗预测场景；局限性则体现为对数据分布外的波动突变适应性不足，这也是传统 ARIMA 模型依赖平稳性假设的固有缺陷。若需提升测试集预测精度，可考虑引入季节项的 SARIMA 模型、融合外部特征的混合模型，或非线性时间序列模型（如 Transformer）进行优化。

## 4.2 Transformer-CNN 混合模型原理

Transformer-CNN 混合模型结合了卷积神经网络 (CNN) 的局部特征提取能力与 Transformer 的全局长依赖建模优势，在时序预测、图像/文本处理等任务中兼顾局部细节与全局关联，成为近年主流的混合建模范式。其核心架构可分为 CNN 特征提取层、Transformer

序列建模层、特征融合层三部分，整体流程如图??所示。

#### 4.2.1 CNN 模块：局部时空特征提取

CNN 通过卷积核的滑动窗口操作捕捉数据的局部关联特征，尤其适用于挖掘时序/空间数据中的局部模式（如电力消耗数据的短期波动、图像的局部纹理）。对于输入序列  $X \in \mathbb{R}^{T \times D}$ （ $T$  为序列长度， $D$  为特征维度），一维卷积层的计算过程为：

$$\mathbf{Z}_i = \sigma \left( \sum_{k=0}^{K-1} \mathbf{W}_k \cdot \mathbf{X}_{i-k:i} + \mathbf{b} \right)$$

其中：-  $K$  为卷积核大小，决定局部特征的感受野；-  $\mathbf{W}_k \in \mathbb{R}^{D \times D'}$  为卷积核权重， $\mathbf{b} \in \mathbb{R}^{D'}$  为偏置项；-  $\sigma(\cdot)$  为激活函数（如 ReLU）， $\mathbf{Z}_i \in \mathbb{R}^{D'}$  为第  $i$  个位置的卷积输出；- 输出特征矩阵  $\mathbf{Z} \in \mathbb{R}^{T \times D'}$  保留了序列的局部时序特征，且通过卷积核的共享权重降低了模型参数量。

为进一步压缩特征维度、提升计算效率，CNN 模块后通常接入池化层（如最大池化/平均池化），公式为：

$$\mathbf{P}_i = \max (\mathbf{Z}_{i:s:(i+1) \cdot s-1})$$

其中  $s$  为池化步长， $\mathbf{P} \in \mathbb{R}^{T' \times D'}$ （ $T' = \lfloor T/s \rfloor$ ）为池化后的特征序列。

#### 4.2.2 Transformer 模块：全局长依赖建模

Transformer 基于自注意力机制（Self-Attention）建模序列的全局依赖，弥补 CNN 仅能捕捉局部特征的不足。其核心是“多头自注意力（Multi-Head Attention, MHA）”，计算过程分为三步：

**步骤 1：Query/Key/Value 投影** 将 CNN 输出的局部特征  $\mathbf{P}$  投影为查询（Query）、键（Key）、值（Value）矩阵：

$$\mathbf{Q} = \mathbf{P}\mathbf{W}^Q, \quad \mathbf{K} = \mathbf{P}\mathbf{W}^K, \quad \mathbf{V} = \mathbf{P}\mathbf{W}^V$$

其中  $\mathbf{W}^Q, \mathbf{W}^K \in \mathbb{R}^{D' \times d_k}$ ， $\mathbf{W}^V \in \mathbb{R}^{D' \times d_v}$  为投影权重， $d_k, d_v$  为注意力维度。

**步骤 2：多头注意力计算** 将  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  分为  $h$  个头部并行计算注意力，单个头部的注意力分数为：

$$\text{Attention}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \text{Softmax} \left( \frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d_k}} \right) \mathbf{V}_h$$



拼接所有头部的输出并投影，得到 MHA 最终结果：

$$\text{MHA}(\mathbf{P}) = \text{Concat}(\text{Head}_1, \dots, \text{Head}_h) \mathbf{W}^O$$

其中  $\mathbf{W}^O \in \mathbb{R}^{h \cdot d_v \times D'}$  为拼接投影权重。

**步骤 3：前馈网络与残差连接**为增强模型表达能力，Transformer 的编码器层包含残差连接和层归一化（Layer Normalization），完整计算为：

$$\begin{aligned} \mathbf{H}_1 &= \text{LayerNorm}(\mathbf{P} + \text{MHA}(\mathbf{P})) \\ \mathbf{H}_2 &= \text{LayerNorm}(\mathbf{H}_1 + \text{FFN}(\mathbf{H}_1)) \end{aligned}$$

其中  $\text{FFN}(\cdot)$  为两层全连接前馈网络， $\mathbf{H}_2 \in \mathbb{R}^{T' \times D'}$  为 Transformer 模块的最终输出。

#### 4.2.3 特征融合与预测层

Transformer-CNN 混合模型的核心是“局部-全局特征融合”，常用融合策略为：1. “拼接融合”：将 CNN 局部特征  $\mathbf{Z}$  与 Transformer 全局特征  $\mathbf{H}_2$  按特征维度拼接，得到融合特征  $\mathbf{F} = \text{Concat}(\mathbf{Z}, \mathbf{H}_2) \in \mathbb{R}^{T \times (D' + D')}$ ；2. “加权融合”：通过可学习权重融合两类特征，公式为  $\mathbf{F} = \alpha \mathbf{Z} + (1 - \alpha) \mathbf{H}_2$  ( $\alpha \in [0, 1]$  为融合权重)。

融合后的特征送入全连接层得到最终预测结果：

$$\hat{y} = \mathbf{F} \mathbf{W}_f + \mathbf{b}_f$$

其中  $\mathbf{W}_f \in \mathbb{R}^{2D' \times 1}$  为预测层权重， $\mathbf{b}_f \in \mathbb{R}^1$  为偏置项， $\hat{y} \in \mathbb{R}^T$  为序列预测值。

#### 4.2.4 模型优势总结

Transformer-CNN 混合模型的核心优势在于：CNN 模块高效捕捉局部时空特征，降低模型对长序列计算的复杂度；Transformer 模块通过自注意力建模全局长依赖，解决 CNN 感受野有限的问题；混合架构兼顾局部细节与全局关联，在电力消耗、交通流量等时序预测任务中，相比纯 CNN 或纯 Transformer 模型具有更优的预测精度。

### 4.3 模型建立

时间序列预测中，针对序列数据的建模主要分为三种核心方式，以区域电力消耗预测为例，各类方式的逻辑与应用场景差异显著。

- 第一种是直接预测序列，本质是自回归建模，仅依赖电力消耗自身的历史时序数据，通过捕捉序列内在的趋势、周期性等自相关特征实现预测，比如利用前 7 天的电力消耗数据预测第 8 天的数值，这种方式建模简单，适用于目标序列自相关性强但缺乏可靠外部数据的场景，不过容易忽略极端天气等外部因素带来的波动。
- 第二种是用其他特征预测序列，核心是依托与电力消耗相关的外部特征构建关联关系，仅通过温度、湿度、风速、diffuse flows 等特征的历史变化规律推导电力消耗的未来值，其优势在于能体现外部因素对目标的因果影响，解释性强，适合电力消耗与外部特征关联紧密但自身时序规律较弱的情况，但会遗漏电力消耗本身的周期性特征。
- 第三种是多序列联合建模，是前两种方式的融合升级，将电力消耗的历史数据与各类外部特征作为多变量时序数据协同建模，既保留目标序列自身的时序惯性，又纳入外部特征的影响，能够捕捉多序列间的交叉依赖关系。

#### 4.3.1 Transformer 单指标序列模型建立

首先,单独使用 Transformer 模型对区域三电力消耗序列数据进行预测,运用前 seq\_len 天的序列，预测下一天的序列数据，得出预测结果和最优参数如下所示：

表 7: Transformer 模型最优超参数组合

seq_len	hidden_dim	num_heads	num_transformer_layers	batch_size	learning_rate
28	64	2	2	64	0.0001

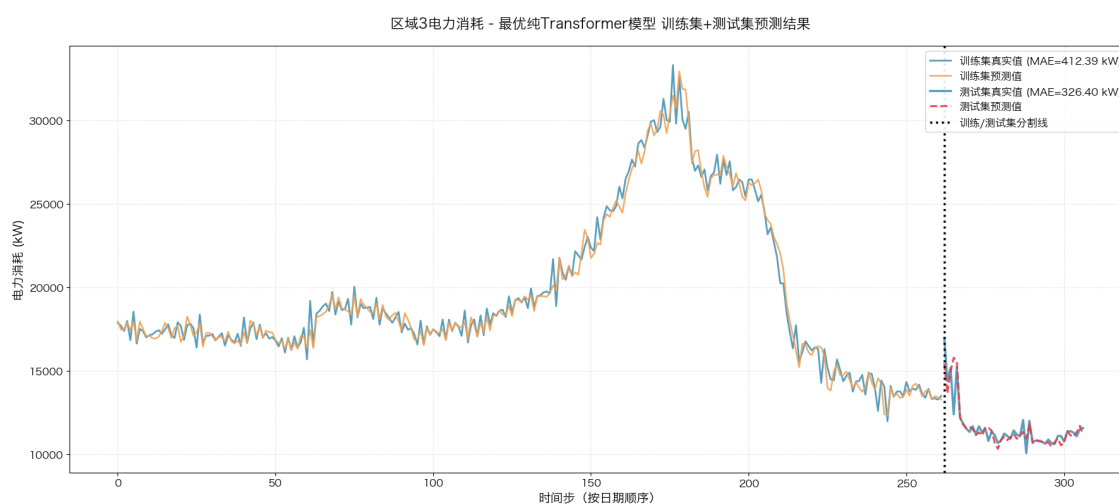


图 5: Transformer 模型单指标序列建模

### 4.3.2 Transformer+CNN 单指标序列模型建立

使用 Transformer+CNN 模型对区域三电力消耗序列数据进行预测，得出预测结果如下图所示：

表 8: Transformer+CNN 模型最优超参数组合

seq_len	hidden_dim	num_heads	num_transformer_layers	kernel_size	batch_size	learning_rate
28	64	2	1	7	64	$5 \times 10^{-5}$

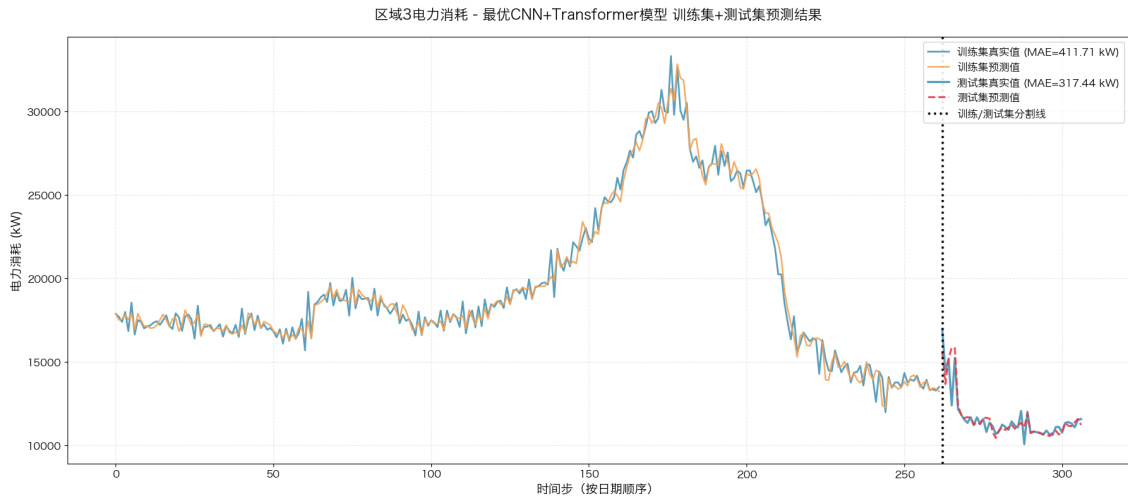


图 6: Transformer+CNN 模型单指标序列建模

从两个模型的预测结果来看：Transformer 模型的测试集预测曲线与真实值偏差相对明显，在真实值快速回落并且后期震荡阶段，预测曲线的跟随性较弱，局部偏离幅度较大；而 Transformer+CNN 模型的测试集预测曲线可更精准地匹配真实值的波动节奏，无论是峰值后的回落趋势还是后期的小幅震荡，均与真实值保持了更高的贴合度。

从 MAE 来看，Transformer+CNN 模型的训练集、测试集 MAE 均低于纯 Transformer 模型，进一步验证了 CNN 局部特征提取 + Transformer 长程依赖捕捉的结构优势——该组合既保留了 Transformer 对时序全局规律的学习能力，又通过 CNN 弥补了其局部波动特征捕捉的短板，最终实现了更优的预测精度。

综上，Transformer+CNN 混合模型在区域三电力消耗单指标预测任务中，展现出更优的局部特征适配性与测试集泛化能力，其结果可为区域电力消耗的精细化调度提供更可靠的参考依据，下文也会围绕 Transformer+CNN 混合模型进行模型建立。

### 4.3.3 模型预测效果差异成因分析

Transformer 与 Transformer+CNN 混合模型的性能差异，核心源于模型结构与电力消耗数据特性的适配性不同。纯 Transformer 依赖自注意力机制捕捉长程时序趋势，但缺乏局部特征提取能力，难以精准跟踪电力消耗的高频震荡、峰值后快速回落等局部突变特征，导致测试集预测跟随性弱、偏差明显。

Transformer+CNN 模型通过 CNN 局部特征提取 + Transformer 长程依赖捕捉的双层架构，实现了数据特性的精准适配：CNN 借助卷积核滑窗运算，高效捕捉短时序尺度的局部波动与突变模式；Transformer 则在局部特征基础上进一步学习长程关联，避免了局部特征被长程信息淹没。同时，CNN 带来的局部归纳偏置提升了模型泛化能力，使其能更好适配测试集未见过的局部波动模式。

MAE 也印证了这一优势，Transformer+CNN 模型训练集与测试集精度均更优，因此后续建模将聚焦该混合模型展开。

### 4.3.4 多特征联合建模

除了单指标序列建模，可以尝试用其他环境特征作为序列特征，对待拟合序列进行模型拟合。本文选择温、湿度等五个指标，用前 seq\_len 个序列长度的环境数据，预测下一天的序列数据，对区域三的电力消耗情况进行预测，结果如下：

表 9: Transformer+CNN 模型最优超参数组合

seq_len	hidden_dim	num_heads	num_transformer_layers	kernel_size	batch_size	learning_rate
28	64	4	4	3	8	$5 \times 10^{-5}$

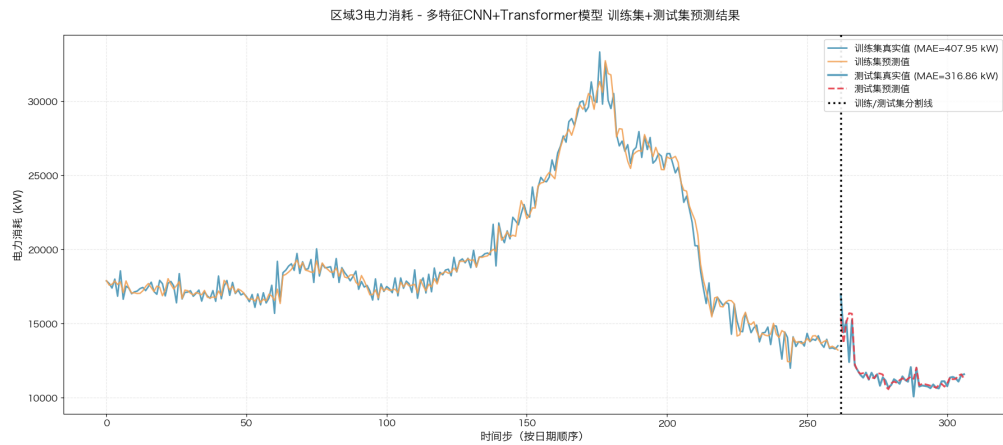


图 7: 多特征联合建模

从预测结果可以看出：测试集泛化能力大幅增强测试集 MAE 降至 316.86kW，测试集预测值与真实值的重合度明显提升——即使在测试集的剧烈波动区间，预测值也能更紧密地跟随真实值的变化，说明外部特征（如温度、风速）有效补充了电力消耗的影响因素，帮助模型更好地解释耗电波动的原因，从而降低了泛化误差。多特征的价值得到验证相比单特征模型，多特征版本在趋势突变、局部剧烈波动”场景下的预测偏差显著缩小，证明温度、湿度等外部特征与电力消耗存在强关联，这些信息的引入让模型从“仅靠自身惯性预测”升级为“结合外部因素的因果性预测”，预测逻辑更合理。

整体来看，多特征 CNN+Transformer 模型的表现更优：训练拟合更精细、测试泛化更稳定，外部特征的加入有效提升了预测精度，是更适合区域 3 电力消耗预测的方案。

#### 4.3.5 多指标联合建模

除了上述两种预测方式，Transformer-CNN 深度学习模型可以同时多个序列数据进行预测，如可以将区域一、区域二、区域三的三个序列数据同时训练，参数共享，除了学习自身特征之外，相互学习彼此的序列特征，预测结果如表 10 和图 8 所示所示：



图 8: 多指标联合建模结果

表 10: Transformer+CNN 模型最优超参数组合

seq_len	hidden_dim	num_heads	num_transformer_layers	kernel_size	batch_size	learning_rate
28	32	16	1	5	32	0.0005

相关预测指标如下表 11 所示：

表 11: 多区域电力消耗预测评估指标

Region	MAE (kW)	MAPE (%)
Zone 1	468.409	1.628
Zone 2	503.571	2.165
Zone 3	343.531	2.864
Average	438.504	2.219

通过三个区域的时序预测曲线可知：测试集泛化方面，预测曲线可紧密跟随真实值的趋势波动（包括区域 1 的平稳变化、区域 2 的局部震荡），验证了“多区域联合建模”对跨区域关联规律的有效捕捉，模型对未观测数据的适配能力优异。各区域及整体的误差指标处于行业内优秀区间：区域维度：区域 1 的预测精度最优（MAPE=1.62%），区域 3 的绝对误差最低（MAE=343.53 kW），区域 2 的误差（MAE=503.57 kW、MAPE=2.16%）也处于合理范围；整体维度：三个区域的平均 MAE 为 438.50 kW、平均 MAPE 仅 2.22%，低误差水平保障了预测结果的可靠性。

## 5 模型对比以及未来展望

### 5.1 模型对比

本研究针对丹吉尔区域电力消耗预测任务，构建了传统线性模型、深度学习单模型及混合模型等多类预测方案，涵盖 ARIMA、纯 Transformer、Transformer-CNN 混合模型（含单指标、多特征、多区域三种建模方式）。各模型在结构设计、特征捕捉能力、预测精度及场景适配性上呈现显著差异，具体对比结果如表 12 所示：

从量化指标与预测效果来看，模型性能呈现明显的层级优势：

- 传统线性模型与深度学习模型的差距：ARIMA 模型因依赖平稳性假设，面对电力消耗的非平稳突变特征时泛化能力显著不足，测试集 MAPE 达 13.89%，是深度学习模型的 5 倍以上；而基于自注意力机制或混合架构的深度学习模型，通过对非线性特征

表 12: 各预测模型核心性能与特性对比

模型类型	核心结构	核心优势	适配场景
ARIMA(2,1,2)	线性自回归 + 差分 + 移动平均	结构简单、样本内拟合稳定	趋势稳定、波动单一的短期预测
纯 Transformer（单指标）	自注意力机制 + 长程依赖建模	高效捕捉长程时序趋势	长程趋势显著、局部波动平缓的数据
Transformer-CNN（单指标）	CNN 局部特征 + Transformer 长程依赖	兼顾局部波动与全局趋势	兼具长程趋势与局部波动的时序数据
Transformer-CNN（多特征）	多环境特征 + “局部-全局” 双层提取	融合外部驱动因素，泛化能力强	外部特征与目标序列强关联的场景
Transformer-CNN（多区域）	跨区域参数共享 + 协同学习	捕捉跨区域关联规律，整体精度优	多区域能耗存在协同关系的场景

的捕捉，大幅降低了预测误差。

- 单模型与混合模型的差距：纯 Transformer 模型虽能学习长程时序规律，但缺乏局部特征提取能力，在真实值“快速回落 + 后期震荡”阶段跟随性较弱；Transformer-CNN 混合模型通过 CNN 补充局部特征提取模块，实现了“局部波动精准捕捉 + 长程趋势有效跟踪”的双重优势，预测曲线与真实值贴合度显著提升。
- 建模方式的优化效果：Transformer-CNN 模型的性能随建模方式升级逐步提升——多特征建模通过融入温度、湿度等外部驱动因素，使测试集 MAE 降至 316.86 kW；多区域联合建模则通过挖掘跨区域能耗协同关系，实现了三个区域平均 MAPE 仅 2.22% 的优异表现，验证了多维度信息融合的价值。

5.2 模型未来展望

- 数据维度优化：丰富社会经济、用电行为等特征，扩充多年度数据，强化模型对能耗驱动因素的刻画；
- 模型结构升级：引入注意力驱动的动态特征融合模块，通过知识蒸馏实现模型轻量化，适配实时调度需求；
- 应用场景延伸：将模型推广至不同气候、产业结构的城市，通过迁移学习提升适配性，与能源调度系统深度对接。