

**Problem Chosen**

**选题 (2)**

2025  
**MCM/ICM**  
Summary Sheet

**Team Control Number**

**231820002**

---

## 干旱条件下植物群落响应的时间序列建模

**231820002 陶德荣**

### 摘要

干旱是全球气候变化背景下对陆地生态系统最重要且最持久的胁迫之一。植被指数 NDVI (Normalized Difference Vegetation Index) 作为反映植被生长状况与地表初级生产力的核心遥感指标，已经广泛应用于评估生态系统对气候异常和环境变化的响应[1]。本研究以美国南部大平原干旱敏感草原区为对象，综合利用 MODIS NDVI 数据集[2]、CHIRPS 降水数据集[3]以及 ERA5-Land 相关气象与土壤湿度数据集，构建了 2000–2024 年 8 日时间尺度的 NDVI–气象多源融合时间序列。通过时间统一、窗口聚合、缺失值插补和标准化等步骤，构建了适用于本次深度学习建模的多变量时间序列数据集。

本文在统一的数据框架下，针对传统的统计模型和多种深度学习模型进行了系统化地对比。传统方法包括持久性模型 (Persistence)、气候平均模型 (Climatology)、多元线性回归与岭回归，以及基于 Box-Jenkins 思想的 ARIMA/SARIMA 时间序列模型。深度学习模型方面，构建了多变量 LSTM 单步预测模型、LSTM Encoder–Decoder (Seq2Seq)、Temporal Convolutional Network (TCN)、Transformer Encoder 和 Temporal Fusion Transformer (TFT) 等结构。所有的模型在实验时都采用了统一的滑动窗口输入形式，每个模型都用了过去的 12 个 8 日时间步的多源特征来预测下一步的 NDVI，在相同的训练/验证/测试划分下，通过 MAE、RMSE、F1-score 与 ROC-AUC 等指标进行不同模型的评估对比。

实验结果表明，多变量 LSTM 单步预测模型在本研究的数据集上具有最佳的整体表现，在预测误差与干旱相关的低 NDVI 状态识别能力方面都是处在领先的水平。LSTM Encoder–Decoder、岭回归与普通线性回归紧随其后，说明在多源特征充足的前提下，适当设计的线性模型仍然具有较强基线能力。TFT 与 Transformer Encoder 在综合指标上也取得了比较好的表现，但是在单步预测精度上略逊于 LSTM 模型。相比之下，ARIMA/SARIMA 等单变量线性时间序列模型难以刻画 NDVI 与多源气象–水文因子的非线性耦合关系，在干旱年份的 NDVI 低谷及快速变化阶段预测偏差较大。本研究结果表明，在干旱背景的植被动态预测任务中，基于多源驱动的 LSTM 模型更适合本类型数据结构，能够为区域干旱监测与生态风险评估提供可靠技术路径。

**关键词:** NDVI，多源数据融合，干旱，时间序列预测，LSTM，ARIMA，深度学习

# 目录

1	介绍	3
1.1	其他假设	4
2	问题分析	4
3	数据预处理与统一建模框架	4
3.1	数据预处理与多源融合	4
3.2	统一的滑动窗口结构	5
3.3	统一评估指标与实验设置	5
4	模型构建	6
4.1	传统模型	6
4.2	深度学习模型	7
5	模型结果	7
5.1	持久性与气候平均模型	10
5.2	线性与岭回归模型	11
5.3	ARIMA 与 SARIMA 模型	12
5.4	LSTM 与 LSTM Encoder–Decoder 模型	13
5.5	TCN、Transformer Encoder 与 TFT 模型	13
6	模型验证	14
7	结论	15
8	总结	16
9	模型评价与优缺点	16
9.1	模型综合评价	16
9.2	模型体系的优点与不足	17
Appendice		19
Appendix A	GitHub代码仓库链接	20
Appendix B	Hugging Face数据集仓库链接	20
Appendix C	Wandb数据集仓库链接	20
Appendix D	可视化仪表盘	20

# 1 介绍

植被对干旱的响应是全球变化生态学中的核心科学问题之一。在气候变暖和极端事件发生频率不断上升的背景下，长期或阶段性干旱会显著影响陆地生态系统的生产力与稳定性。归一化差值植被指数 NDVI (Normalized Difference Vegetation Index) 是基于红光与近红外波段反射率构建的简单光谱指数，能够较好地反映出地表绿色植被覆盖度和光合活性 [1]。由于其连续观测、覆盖范围广和可比性强等特点，NDVI 已经被广泛用来监测植被生长季节节律、土地退化、生物量变化以及生态系统对气候与人类活动影响的响应 [1]。

大量的研究表明，植被对干旱的响应不仅仅取决于降水的即时变化，还会受到干旱时期的持续时间与植被累积水分亏缺的影响，不同的生物群系对于干旱时间尺度的敏感性存在显著的差异 [2]。在干旱敏感地区，如美国南部大平原，降水年际变率较大，极端干旱事件时有发生，草原和农牧交错带的植被对于降水和土壤水分的波动具有很强的敏感性，这些地区的NDVI 在干旱发生时就会出现明显的下降，并在干旱结束后出现一定的滞后性恢复 [2]。因此，基于多源气象-生态数据构建出能够较为准确地刻画出 NDVI 的时滞响应与干旱过程的时间序列预测模型，对于区域干旱监测、生态风险评估与适应性管理具有重要现实意义。

在本研究采用的数据上，卫星遥感和再分析的产品为干旱与植被响应研究提供了丰富的数据源。MODIS 植被指数产品（如 MOD13 系列）可以提供长期连续的全球 NDVI 序列，这些数据可以用于监测植被活动在季节与年际尺度上的变化 [4]。CHIRPS (Climate Hazards Group InfraRed Precipitation with Stations) 降水数据集则是通过融合红外卫星观测与地面站点，实现了  $0.05^{\circ}$  空间分辨率的多年降水格点数据产品，尤其适用于趋势分析和干旱监测 [3]。结合 ERA5-Land 等再分析数据集，可以同时获得土壤湿度、气温和潜在蒸散等其他的气象要素，这些数据集为本研究构建多源驱动的 NDVI 模型提供了基础。

在本研究使用的参考的方法上，传统时间序列模型（如 ARIMA/SARIMA）已经广泛应用于单变量气候与水文序列的预测，能够刻画线性自相关结构和季节性成分。然而，NDVI 的变化往往是多种环境因子共同作用的结果，具有非线性和时间滞后的特征，仅仅依赖 NDVI 自身的过去值是难以去精确刻画干旱背景下的动态过程。近些年来，深度学习模型，尤其是长短期记忆网络（LSTM）、卷积时序网络（TCN）以及基于注意力机制的 Transformer 模型，在复杂时间序列预测任务中展现出强大能力 [6]。这类模型可以在统一框架下利用多源输入，自动学习长程依赖和非线性映射关系，逐渐成为气候与生态时间序列建模的重要方向。

基于以上的数据和方法，本文将以美国南部大平原干旱敏感草原为研究对象，基于 2000–2024 年的多源遥感与气象数据，构建 8 日尺度的 NDVI–气象多变量时间序列。我们在统一的数据预处理和滑动窗口结构下，系统化地比较了传统统计模型（Persistence、Climatology、线性回归、岭回归、ARIMA/SARIMA）与多种深度学习模型（LSTM、LSTM Encoder–Decoder、TCN、Transformer Encoder、TFT）的表现。通过 MAE、RMSE 等回归指标以及 F1-score、ROC-AUC 等干旱识别指标，评估了不同模型在干旱条件下对于 NDVI 预测任务中的表现的优劣，并重点讨论 LSTM 与 ARIMA/SARIMA 在本类多源非线性数据上的适用性差异。研究目标是构建一套具有可推广性的多源时间序列建模框架，为未来在不同区域开展干旱监测与植被响应研究提供方法参考。



图1. 选取研究区域

## 1.1 其他假设

为了在有限的数据与计算资源条件下构建出可行并且具有一定的解释力的模型体系，本文在数据处理与模型构建过程中作出了以下一些合理的假设：

首先，在数据集的可靠性上，本研究假设经过质量控制处理后的 MODIS NDVI 产品能够真实地反映出研究区的地表绿色植被覆盖度的时空变化，尽管在极端天气或者观测条件不佳的时候可能存在一定的局部误差。同时，还假设了 CHIRPS 降水数据和 ERA5-Land 再分析产品在区域平均尺度上的系统误差相对较小，足以支撑时间序列建模和干旱监测的需求。

其次，在空间表示方面，假设研究区内的植被类型比较一致，都是以草原和农牧混合生态系统为主的植被，区域平均的 NDVI 可以近似地看作该区的整体生态状态。虽然实际地表存在一定的不一致情况，但是在本研究中将区域平均 NDVI 作为因变量有助于简化模型结构与计算复杂度。

第三，在时间尺度处理上，假设本研究以 NDVI 的 8 日时间窗口作为主时间轴是合理的，即日尺度气象要素通过 8 日窗口聚合为累积降水、平均土壤湿度、平均气温与累积潜在蒸散后，仍然能够保留对 NDVI 变化最关键的气候信号，不会因为过度平滑处理而过滤掉干旱过程的演变特征。

第四，在干旱识别标准上，假设可以通过训练集 NDVI 分布的低分位数（如 20% 分位）来划分“相对低 NDVI 状态”和“正常及偏高状态”，本研究将基于这样的二元划分而计算出的 F1-score 与 ROC-AUC 等指标来近似衡量模型在识别干旱相关低植被状态方面的能力。

最后，在不同模型的可比性方面，假设实验通过统一的滑动窗口长度（12 个 8 日时间步）、输入变量集合、损失函数与评估指标，可以确保传统统计模型和深度学习模型之间的比较是公平的、可比的。

## 2 问题分析

本研究要解决的核心问题，是在干旱敏感区利用多源气象-生态数据来准确地预测出 NDVI 的时间演变，并评估出不同的模型在干旱相关低 NDVI 状态识别方面的能力。为此，本研究分别从时间尺度、影响因子层面对该问题进行分析。

从时间尺度上看，NDVI 序列往往具有显著的年周期性。年周期主要是反映植被生长季节性节律的：植被一般在春季由低向高快速上升，夏季维持较高水平，秋季又会逐步回落到枯黄状态。同时 NDVI 也会受到降水异常、气温偏差以及干旱事件的影响，在极端干旱年份，NDVI 在整个生长季期间都会出现持续偏低的状况。因此，本实验研究的模型需要能够既识别季节尺度的周期变化，又对干旱过程的累积效应保持敏感。

从影响因子上看，降水、土壤湿度、气温和潜在蒸散等变量共同决定了植被水分状况和生长条件。降水是最直接的水源输入，土壤湿度则对短时的降水波动存在一定的平滑和滞后作用。气温和潜在蒸散则决定了大气对水分的“需求”。在干旱过程中，这些变量之间并非简单线性关系，而是通过复杂的能量-水分平衡过程耦合。对于 NDVI 的影响也会表现为非线性的、带阈值的响应：降水减少到一定程度后，NDVI 才会出现明显下降，之后在土壤水分恢复前维持较低水平。

## 3 数据预处理与统一建模框架

### 3.1 数据预处理与多源融合

首先，本研究对于 NDVI、降水、土壤湿度、2 m 气温和潜在蒸散等数据进行了统一地读取。NDVI 使用了 MODIS 植被指数产品（MOD13 系列），时间分辨率为 8 日合成，经过质量控制后获得了长期连续的植被活动序列。降水则是采用了 CHIRPS 数据集，该数据集的空间分辨率约为  $0.05^{\circ}$ ，时间分辨率为日尺度，比较适合本实验的用于干旱监测与趋势分析。ERA5-Land 则是提供了与陆地相关的大气和土地表面变量，包括土壤湿度、气温与潜在蒸散等，在再分析框架下跟其他数据集具有一致性。

在时间对齐方面，本研究将 NDVI 的 8 日时间步作为主轴，对于其他的日尺度的变量进行了聚合操作。具体做法是：对于每个 NDVI 时间点，向前追溯该时间点对应的 8 日窗口，将降水和潜在蒸散按 8 日累积求和，将土壤湿度和气温按 8 日平均求值，从而得到了和已有的 NDVI 数据一一对应的多源样本。这样在实验时既能够保留 NDVI 合成周期的信息，又能够将高频的气象变量转化为和植被响应时间尺度相互匹配的影响变量。对于数据集中的出现的少量的缺失值，在实验中会在时间维度上采用线性插值和邻近插值进行对应的修复，而对那些无法进行有效插补的极端异常点则会进行剔除。

在特征工程上，为了能够帮助模型更好地识别出季节性的变化，本实验引入了年内日序（Day of Year, DOY）并对该指标进行了正余弦编码，将  $\sin(2\pi DOY/365)$  和  $\cos(2\pi DOY/365)$  作为两个附加特征添加进了输入属性中，使得模型可以通过连续的角度空间表达时间位置。这种方法在时间序列建模中已被广泛用于增强模型对周期性信号的识别能力。随后，本实验对于所有的输入特征都采用了标准化处理，以训练集为基准估计均值和标准差，并将同一变换应用于验证集和测试集，确保不同变量在数值范围上的可比性。

在时间和数据集的划分上，本实验按照时间顺序将 2000–2014 年作为训练集，2015–2018 年作为验证集，2019–2024 年作为测试集。这样的划分方式避免了未来信息泄漏，并且保证了模型在训练阶段只能看到过去的数据，在测试阶段通过滚动预测检验其泛化能力。验证集用于超参数选择和早停控制，测试集则完全不参与模型拟合的过程，仅仅用于实验最终的性能评估。

## 3.2 统一的滑动窗口结构

为了能够在不同模型之间实现公平的对比，本研究为所有的模型设计了统一的滑动窗口结构。设窗口长度为  $L = 12$ ，该设置表示了模型会使用过去的 12 个 8 日时间步的多源特征来预测第  $L + 1$  个时间步的 NDVI。对原始时间序列  $\{xt, yt\}$  进行滑动截取，构造出样本对  $(Xi, yi)$ ，其中：

$$Xi = [xi - L + 1, xi - L + 2, \dots, xi], \quad yi = yi + 1.$$

这里  $xt$  是包含 NDVI 历史、降水、土壤湿度、气温、潜在蒸散以及时间编码等在内的特征向量。对于深度学习模型而言， $Xi$  的形状为  $L \times F$ ，其中  $F$  为特征维数；对于线性回归等模型，则将  $Xi$  展平为长度为  $L \cdot F$  的向量。

每一个样本的目标值  $yi$  对应于窗口结束后的下一期 NDVI。这样设计的好处是：一方面确保所有模型在使用信息的时间范围上高度一致；另一方面为后续评估提供统一的“对齐索引”，便于将基线模型（如 ARIMA）的预测结果与深度学习模型的输出在相同时间点进行对比。对于 LSTM Encoder–Decoder 等多步预测模型，内部可以输出多个未来时间步的预测值，但在与其他模型做单步对比时，只取其第 1 步预测结果参与统一评估。

在干旱识别相关的二分类评估中，本研究基于训练集 NDVI 的经验分布，将某一低分位数（20% 分位）作为阈值  $\theta$ ，将  $yt < \theta$  的样本标记为“相对低 NDVI 状态”，记作 1，其余样本标记为 0。对于连续预测值  $y^t$ ，通过与阈值  $\theta$  比较即可得到预测标签，进而计算 F1-score。为了衡量模型对不同阈值的整体识别能力，还利用  $-y^t$  作为“干旱风险”得分计算 ROC-AUC，避免因为人为选择单一概率阈值而造成偏差。

## 3.3 统一评估指标与实验设置

在评估指标上，本研究采用了两类不同的评估指标。第一类是回归误差指标，包括了平均绝对误差（MAE）和均方根误差（RMSE），这两个指标是用来度量预测值与真实 NDVI 在数值上的偏差。第二类是基于 NDVI 低分位阈值转换得到的二分类指标，包括了 F1-score 和 ROC-AUC，这两个指标是用来评估模型对于低 NDVI 状态（可视为与干旱相关）的识别能力。

在实验设置方面，所有深度学习模型都采用了相近的训练策略：用均方误差作为损失函数，使用 Adam 优化器进行权重更新，在训练集上进行逐轮的迭代，并通过验证集监控早停条件，防止模型出现过拟合。传统模型则是会按各自的拟合方式估计参数，例如线性回归通过最小二乘法求解权重，ARIMA/SARIMA 通过极大似然估计参数并选取信息准则最优模型阶次。为保证结果的可比性，所有的模型的预测都在同一测试集索引上进行对齐，并统一通过上述指标进行评估。

## 4 模型构建

本研究实现了一系列的传统统计模型与深度学习模型构建。所有的模型都是以长度为 12 的时间窗口作为输入，以第 13 个时间步的 NDVI 作为单步预测的目标。下面对各模型的具体构建方式进行了简要的说明，以下的模型都已经在实际代码进行了实现。

### 4.1 传统模型

传统模型主要作为基线进行参照，用于刻画仅依赖历史 NDVI 或简单统计规律时能达到的预测效果。

#### (1) 持久性模型 (Persistence)

持久性模型直接将上一期 NDVI 作为下一期的预测值，即

$$\hat{y}^t = y_t$$

该模型无需进行训练，仅依赖 NDVI 的短期自相关，是最简单的时间序列基线。

#### (2) 气候平均模型 (Climatology)

气候平均模型将“年内日序 DOY 的多年平均 NDVI”作为预测值。具体的做法是，在训练集上按 DOY 统计多年 NDVI 的平均值，对于给定预测日期，查找对应 DOY 的平均 NDVI，作为当前时刻的预测。这一模型能够反映季节性，但无法区分不同年份的干旱差异。

#### (3) 多元线性回归与岭回归 (Linear / Ridge)

线性类模型的输入是展平后的窗口特征，将 12 个时间步、每步 F 个特征拼接成长为 12F 的向量，通过线性变换预测 NDVI：

$$\hat{y}^t = w^T z + b$$

其中  $z$  是展平后的特征向量。岭回归在此基础上加入 L2 正则化项，用来缓解多重共线性和过拟合问题。

#### (4) ARIMA / SARIMA 单变量时间序列模型

ARIMA 模型采用了  $(p,d,q)$  结构，仅以 NDVI 自身作为建模对象，通过差分、自动回归项与移动平均项刻画线性自相关结构和趋势。SARIMA 在此基础上进一步引入季节部分  $(P,D,Q,s)$ ，试图捕捉 NDVI 的年周期成分。代码中使用 statsmodels 库对训练集 NDVI 进行拟合，通过信息准则 (AIC) 选择最优阶次。最终选出的 ARIMA 为

ARIMA(2,0,2), SARIMA 为 SARIMA(2,0,2,0,0,46)。由于受限于模型形式, 这两类模型只使用 NDVI 单变量, 不纳入降水、土壤湿度等外生驱动因子。

## 4.2 深度学习模型

深度学习模型都是以长度为 12 的多变量序列作为输入, 利用网络结构自动学习时间依赖与变量间的非线性关系。所有模型的训练都采用了均方误差损失函数和 Adam 优化器, 并通过验证集早停控制。

### (1) LSTM 单步预测模型 (LSTM Many-to-One)

该模型使用了单层 LSTM 读取输入序列, 输入张量形状为  $(L,F)$ , 其中  $L=12$ 。LSTM 在时间维度上会逐步更新隐藏状态, 最后一个时间步的隐藏状态  $h_L$  作为整个序列的编码表示。随后通过一层全连接线性层, 将  $h_L$  映射到标量输出  $y^{\wedge}$ , 即第 13 个时间步的 NDVI 预测。

### (2) LSTM Encoder–Decoder (Seq2Seq)

该模型采用了典型的编码器–解码器结构。编码器部分为一层或多层 LSTM, 读取长度为 12 的历史序列, 将其压缩为末状态的隐藏状态与细胞状态。解码器部分以编码器状态作为初始状态, 在训练时采用了教师强制 (teacher forcing), 逐步输出未来多个时间步的 NDVI 预测。本研究的评估中, 为了与其他单步预测模型的评估保持一致, 本实验仅取该多步预测序列中第 1 个时间步的预测值参与对比。

### (3) Temporal Convolutional Network (TCN)

TCN 模型采用了一维卷积网络对时间序列进行建模, 关键特征包括因果卷积和膨胀卷积两个特征。因果卷积保证了模型在计算当前时间步输出时, 只使用当前及之前的时间步信息; 而膨胀卷积则是通过在卷积核中跳过一定间隔的时间步, 显著扩大感受野, 从而使得模型能够在较浅的层数下捕捉长程依赖。

### (4) Transformer Encoder 时间序列模型

该模型基于 Transformer Encoder 架构, 通过多头自注意力机制学习不同时间步之间的依赖关系。输入序列首先经过了线性变换和位置编码的处理, 将时间信息注入特征表示, 随后通过多层的 Encoder Block 对数据进行了特征提取。每个 Encoder Block 包含多头自注意力子层和前馈网络子层, 并且配合残差连接和层归一化。最终对编码后的序列进行池化, 最后模型会通过全连接层输出标量 NDVI 的预测结果。

### (5) Temporal Fusion Transformer (TFT)

TFT 模型在 Transformer 的基础上引入了变量选择网络、门控残差网络以及多时间尺度的信息融合模块, 用于多步时间序列预测 [9]。在本研究中, TFT 被配置为了单步预测任务, 输入为多变量历史序列与部分时间相关特征, 模型通过静态特征编码、历史路径、未来路径与注意力层综合提取信息。

## 5 模型结果

在上述统一的数据处理和建模框架下, 该部分对所有模型在测试集上的表现进行了评估。评估指标包括 MAE、RMSE、F1-score 和 ROC-AUC, 分别反映数值拟合误差与对低 NDVI 状态 (与干旱相关) 的识别能力。整体对比结果综合可视化图如下:

### test\_MAE

SARIMA(2, 0, 2, 0, 0, 0, 46)\_Baseline

ARIMA(2, 0, 2)\_Baseline

Ridge\_Regression\_Baseline

Linear\_Regression\_Baseline

Climatology\_Baseline

Persistence\_Baseline

TransformerEncoder\_TS\_dm64\_h4\_L2

TFT\_hid64\_heads4

TCN\_channels(64, 64, 64)\_k3

LSTM\_ManyToOne\_hid64\_L1

图2. 测试集平均绝对误差 (test\_MAE)

### test\_RMSE

SARIMA(2, 0, 2, 0, 0, 0, 46)\_Baseline

ARIMA(2, 0, 2)\_Baseline

Ridge\_Regression\_Baseline

Linear\_Regression\_Baseline

Climatology\_Baseline

Persistence\_Baseline

TransformerEncoder\_TS\_dm64\_h4\_L2

TFT\_hid64\_heads4

TCN\_channels(64, 64, 64)\_k3

LSTM\_ManyToOne\_hid64\_L1

图3. 测试集均方根误差 (test\_RMSE)

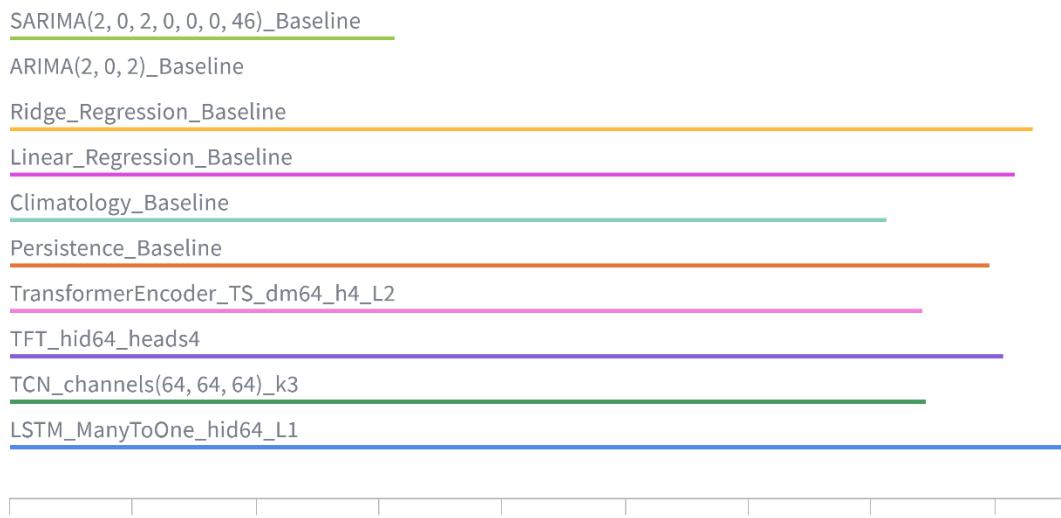
**test\_F1**

图4. F1-score 对比 (test\_F1)

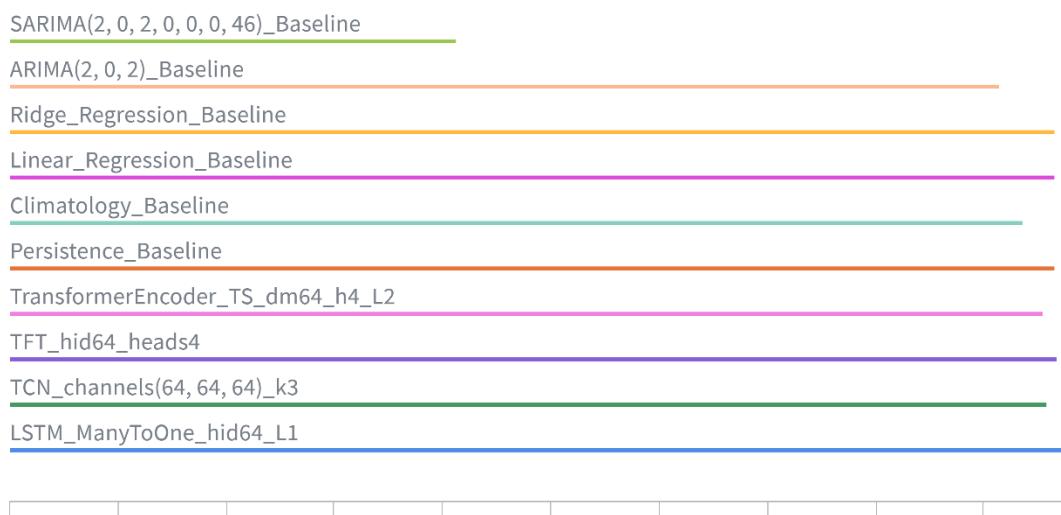
**test\_ROC\_AUC**

图5. ROC-AUC 对比 (test\_ROC\_AUC)

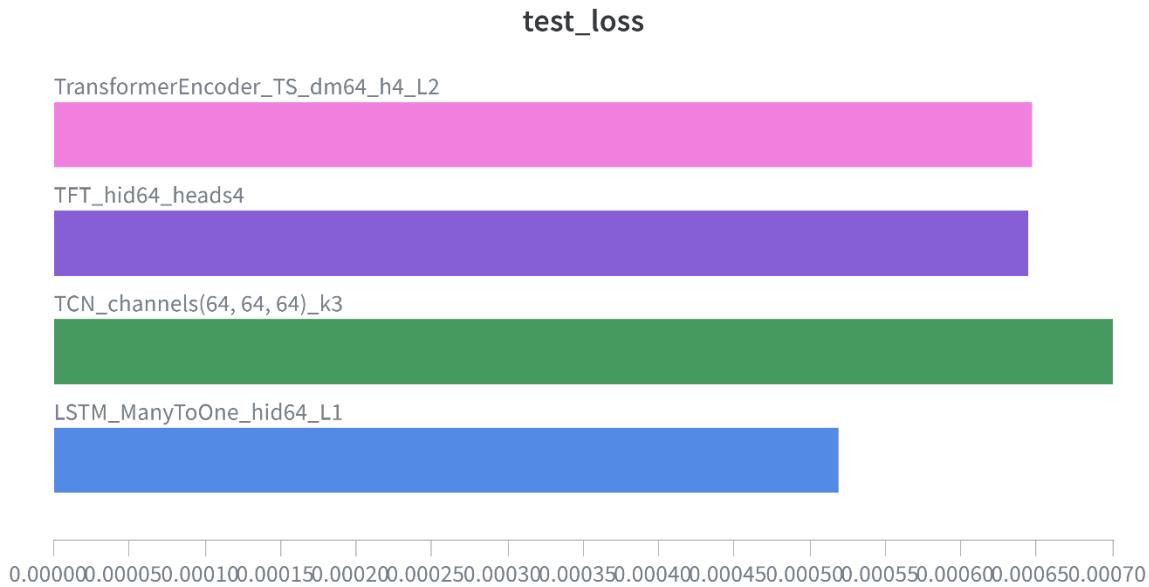


图6. 测试集损失曲线 (test\_loss)

从整体的评估来看，多变量 LSTM 单步预测模型是所有模型中表现最好的，它的测试集 MAE 约为 0.015，RMSE 约为 0.023，F1-score 约为 0.86，ROC-AUC 接近 0.98。岭回归和普通线性回归在多源特征输入下表现也很优秀，MAE 分别约为 0.0158 和 0.0162，F1-score 分别约为 0.83 和 0.82，这说明线性模型在特征工程得当的情况下仍然可以具备较强的基线预测能力。LSTM Encoder-Decoder 模型在第 1 步预测上的 MAE 约为 0.0169，RMSE 约为 0.0242，在误差水平上略高于 LSTM 单步模型，但在 F1-score 和 ROC-AUC 上保持了接近的水平。

在深度学习模型中，TFT 和 Transformer Encoder 的 MAE 大约在 0.018–0.018 前后，RMSE 约为 0.025–0.026，F1-score 和 ROC-AUC 也处于较高水平，说明注意力机制模型在多源时间序列预测中具备良好的整体性能。TCN 的 MAE 约为 0.019，略微高于前述模型，但是依然显著优于基于单变量的 ARIMA/SARIMA 和简单基准模型。Climatology、ARIMA 和 SARIMA 的误差整体较大，其中 ARIMA 和 SARIMA 的误差远高于其他模型，表明这两个模型在当前数据设定下难以有效刻画 NDVI 的复杂时序结构。

下面结合各模型的预测 – 真实对比图和残差图，对各模型的表现进行更加具体的分析。

## 5.1 持久性与气候平均模型

持久性模型（Persistence）和气候平均模型（Climatology）通过简单规则给出预测，这两个模型的表现为后续模型提供了“下限”和“季节参考”。从持久性模型的预测 – 真实对比图可以看到：

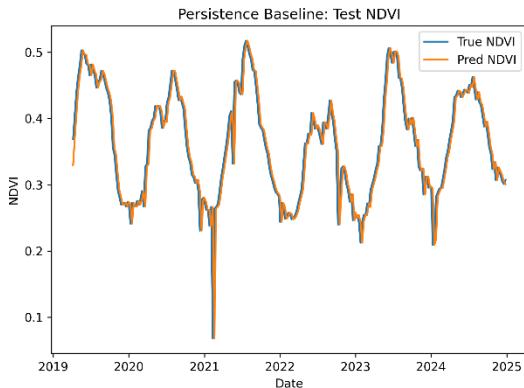


图7. persistence\_pred\_vs\_true

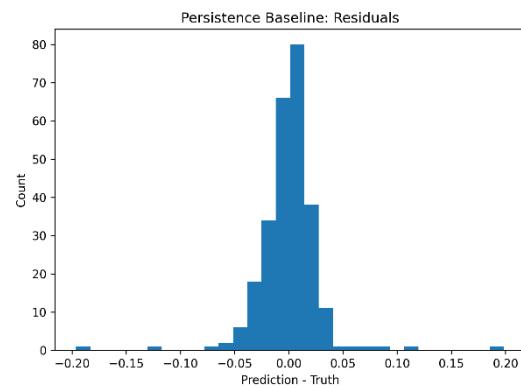


图8. persistence\_residuals

持久性预测与真实值非常相近，近乎完美的预测了所有的真实值，说明该模型针对这个数据集的效果非常好。后续不再用这个模型进行比较。

气候平均模型的预测 - 真实对比与残差图为：

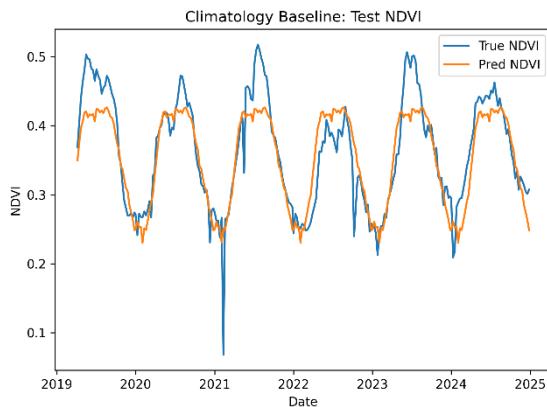


图9. climatology\_pred\_vs\_true

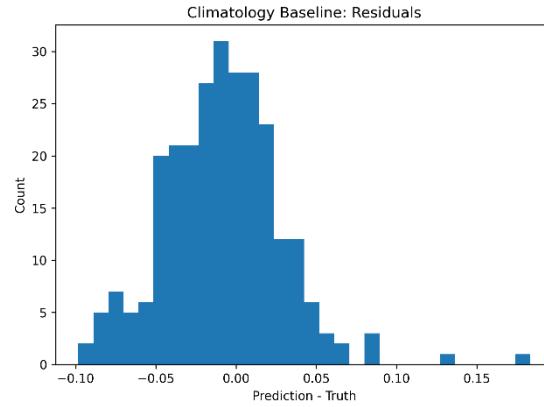


图10. climatology\_residuals

该模型能够比较好地复现出年周期特征，但是在干旱的年份，尤其是生长季中期持续干旱时，预测出的 NDVI 明显偏高，残差呈现持续的负偏差。这表明仅仅基于多年平均气候状态无法准确反映年际间的干旱差异。

## 5.2 线性与岭回归模型

线性回归与岭回归使用多源特征展平后的向量作为输入，能够利用到降水、土壤湿度和温度等变量来进行预测。其预测 - 真实对比和残差图如下：

线性回归：

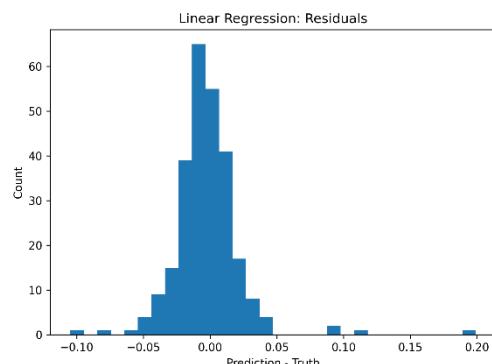
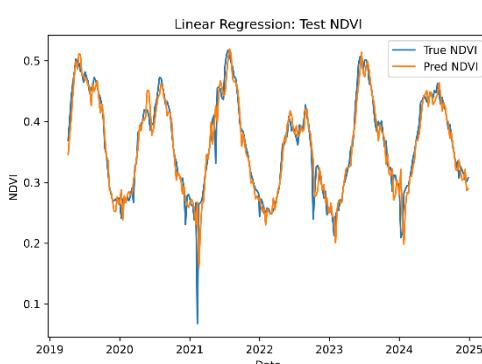


图11. linear\_pred\_vs\_true

岭回归:

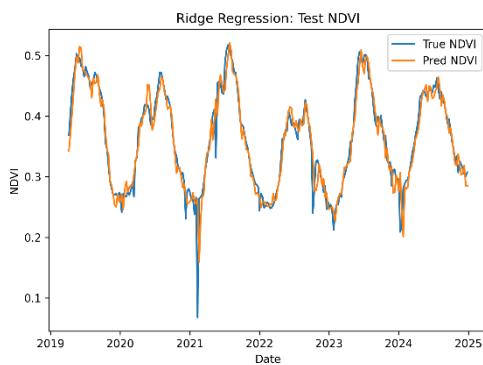


图12. linear\_residuals

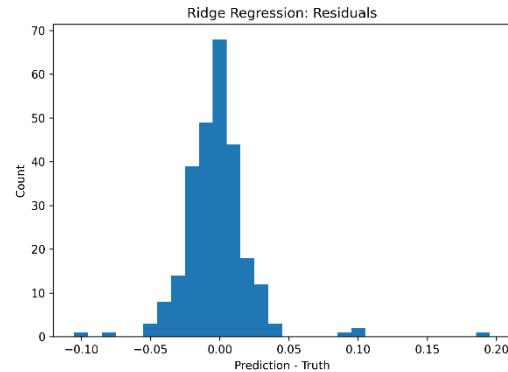


图13. ridge\_pred\_vs\_true

图14. ridge\_residuals

从图中可以看出，线性模型在整体趋势上比气候平均模型表现得更好，能够在一定程度上跟踪 NDVI 的年际波动和干旱年份的偏低状态。岭回归的残差分布比普通线性回归更为集中，说明正则化有助于稳定模型参数，减小训练噪声对预测的影响。

### 5.3 ARIMA 与 SARIMA 模型

ARIMA 和 SARIMA 模型仅仅是基于 NDVI 单变量进行的拟合，这两个模型的预测 - 真实对比和残差图如下：

ARIMA(2,0,2):

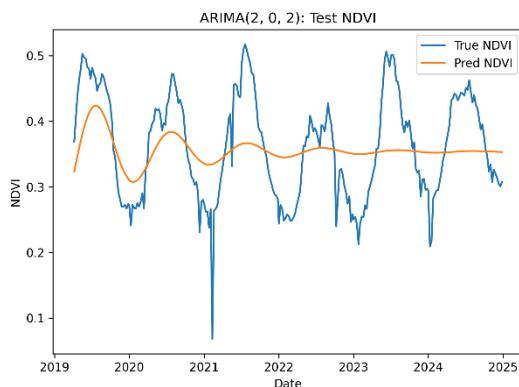


图15. ARIMA\_pred\_vs\_true

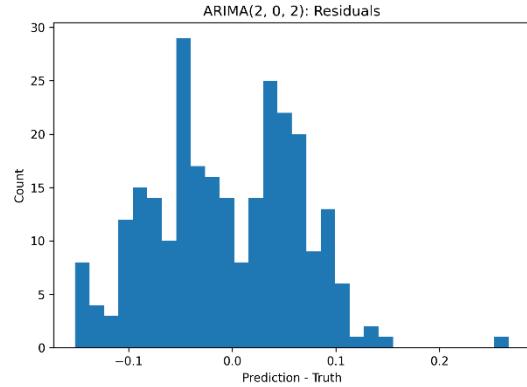


图16. ARIMA\_residuals

SARIMA(2,0,2,0,0,0,46):

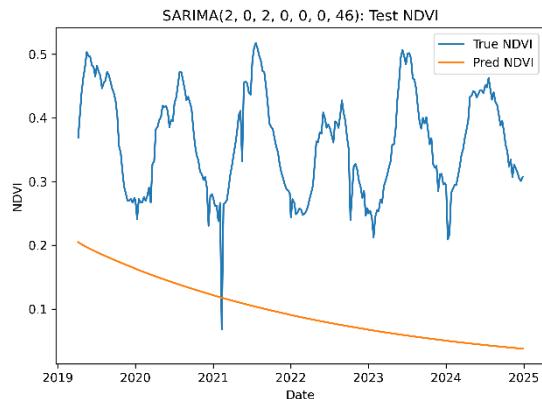


图17. SARIMA\_pred\_vs\_true

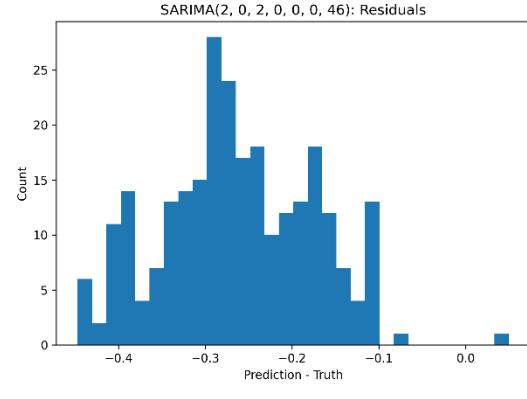


图18. SARIMA\_residuals

从数值指标来看，ARIMA 的 MAE 和 RMSE 明显高于其他的模型，F1-score 接近 0，说明在采用 NDVI 低分位数阈值进行干旱识别时，该模型几乎无法正确识别出低 NDVI 状态。SARIMA 在引入季节项后，整体误差不降反升，测试集 MAE 和 RMSE 在所有模型中是最为突出的，ROC-AUC 也是明显偏低的情况。而从残差图来看，两者在干旱年份生长季中期存在持续的偏差，这就表明了仅仅依靠 NDVI 的自相关和固定季节结构难以应对真实气候条件下的复杂干旱过程。

## 5.4 LSTM 与 LSTM Encoder-Decoder 模型

多变量 LSTM 单步预测模型（LSTM Many-to-One）的损失曲线与预测 - 真实对比如下：

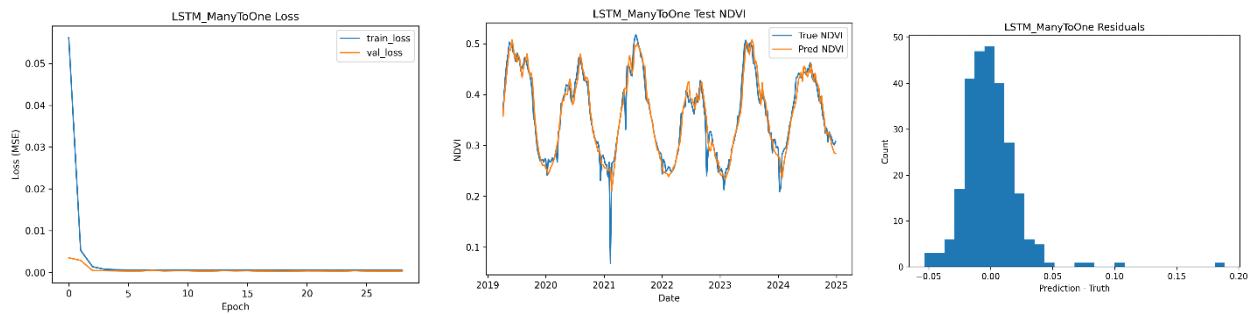


图19. LSTM\_ManyToOne\_loss 图20. LSTM\_ManyToOne\_pred\_vs\_true 图21. LSTM\_ManyToOne\_residuals

可以看出来，该模型在训练和验证阶段的损失曲线收敛比较稳定，测试集上能够比较准确地跟踪 NDVI 的季节变化与干旱年份的偏低区间。残差分布整体集中在零附近，未见明显的系统性偏差，说明该模型对整体趋势与异常事件都有着比较好的刻画。

LSTM Encoder-Decoder 的第 1 步预测结果表现略逊于 LSTM 单步模型，但是整体表现还是比较好的。其相关图如下：

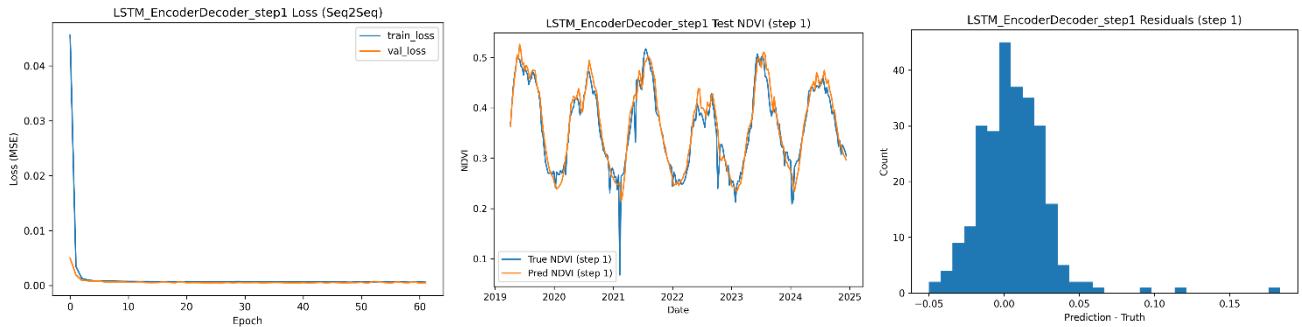


图22. lstm\_encdec\_loss 图23. lstm\_encdec\_pred\_vs\_true\_step1 图24. lstm\_encdec\_residuals\_step1

由于该模型原本设计用于多步预测，在只考察第 1 步时，其优势未能完全体现。尽管如此，误差水平与 F1-score、ROC-AUC 仍然说明它在刻画 NDVI 与多源气象因子关系方面具备较强能力。

## 5.5 TCN、Transformer Encoder 与 TFT 模型

TCN 模型的损失与预测 - 真实对比如下：

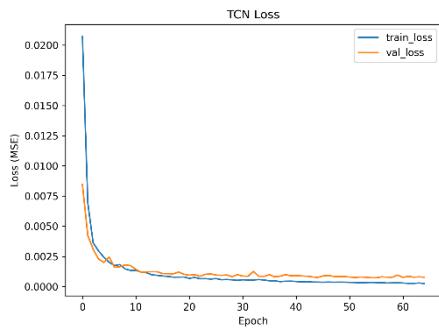


图25. tcn\_loss

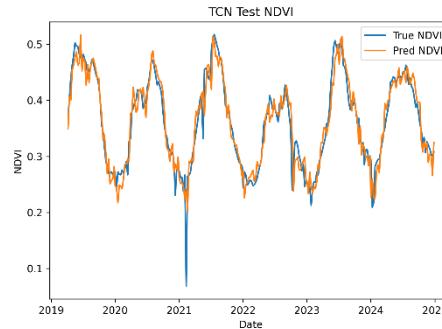


图26. tcn\_pred\_vs\_true

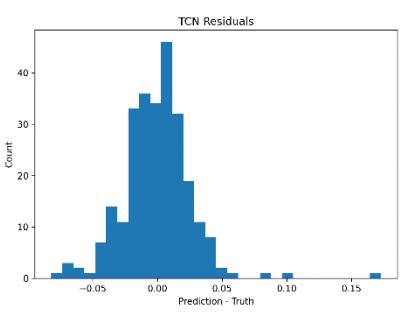


图27. tcn\_residuals

TCN 在数值误差上略高于 LSTM 系列，但是仍然明显优于 ARIMA/SARIMA 这两个模型。该模型在 NDVI 快速变化的阶段比持久性模型更加敏感，但是在一些极端低值上存在一定的高估。

Transformer Encoder 模型也表现出较好的综合能力：

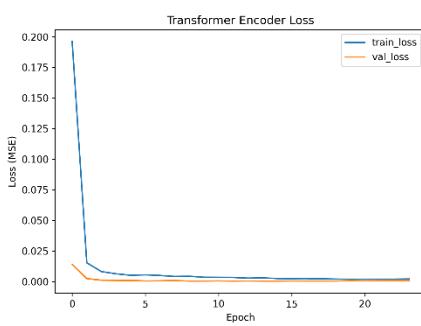


图28. transformer\_loss

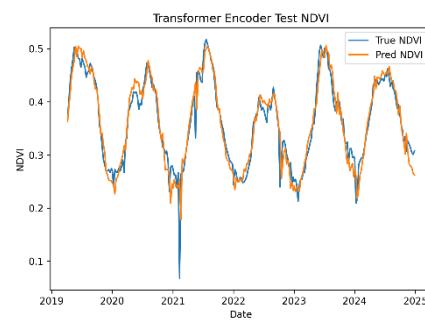


图29. transformer\_pred\_vs\_true

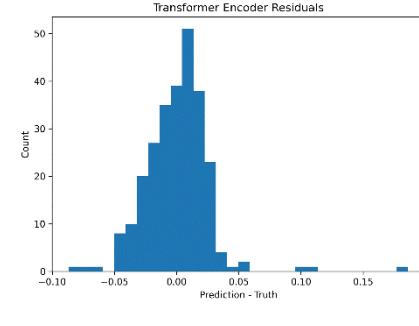


图30. transformer\_residuals

该模型的 MAE、RMSE 与线性模型接近，ROC-AUC 接近 0.95 以上，说明该模型能够通过注意力机制在一定程度上识别出关键的时间步与变量，但是在当前数据规模与任务设置下，还是没有超越 LSTM 模型。

TFT 模型的结果如下：

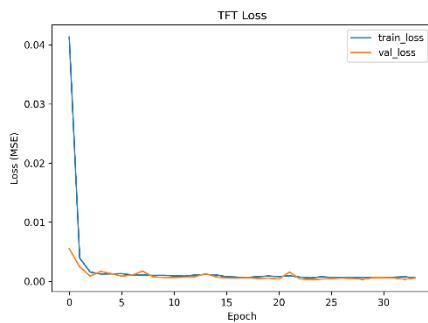


图31. tft\_loss

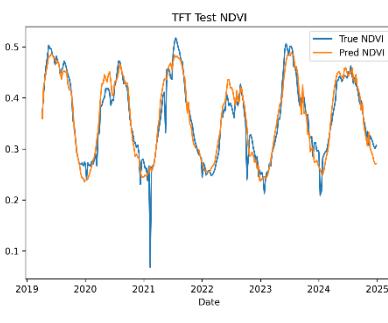


图32. tft\_pred\_vs\_true

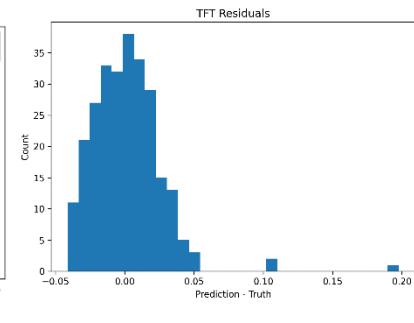


图33. tft\_residuals

TFT 的误差略高于 LSTM 单步模型，但是 ROC-AUC 略高于线性与 Ridge 模型，说明这个模型在低 NDVI 状态识别方面具有一定的优势。由于 TFT 结构更复杂，对数据量与调参较为敏感，在本研究数据规模下，这个模型的潜力还没有完全发挥。

## 6 模型验证

本实验的模型验证会从误差结构、干旱识别能力、残差分析以及训练稳定性等方面进行。

首先，从回归误差指标来看，多变量 LSTM 单步模型的 MAE 和 RMSE 都是全部模型中最小的，岭回归、线性回归与 LSTM Encoder–Decoder 紧随其后。TFT 和 Transformer Encoder 虽然有一定的劣势，但是整体的表现仍然处在一个较优的水平，TCN 则是稍逊一筹。Climatology、ARIMA 和 SARIMA 的误差明显比较大，其中 SARIMA 的表现最差。

从干旱相关的二分类指标上看，使用训练集 NDVI 低分位数阈值划分的低 NDVI 状态作为“干旱近似指标”时，多变量 LSTM 单步模型在 F1-score 和 ROC-AUC 上都取得了最高或接近最高的数值。岭回归、线性回归与 LSTM Encoder–Decoder 在 F1-score 上也具有较好表现，说明这些模型能够比较好地区分出低 NDVI 与正常状态。TFT 和 Transformer Encoder 的 ROC-AUC 指标接近甚至略高于线性模型，这样的结果表明了注意力机制模型在排序意义上的识别能力比较强。相比之下，ARIMA 在 F1-score 上接近 0，几乎无法识别低 NDVI 状态，而 SARIMA 的 ROC-AUC 也是明显低于其他模型的，这说明了该模型在本实验的二分类任务下适用性是比较差的。

从残差分析角度来看，多变量 LSTM 模型的残差在时间上分布是比较均匀的，没有出现明显的季节性偏差情况，也没有在干旱年份集中出现系统性地高估或者低估。岭回归与线性回归的残差分布则是呈现出了轻微的异方差性情况，但是模型整体的偏差是在可控范围内的。TCN、TFT 与 Transformer 的残差在极端低 NDVI 时间段有一定的高估倾向，但是在多数时间段与 LSTM 模型相近。相反，ARIMA 和 SARIMA 的残差在部分年份的生长季中表现出明显的持续偏差，这和它们仅仅依赖 NDVI 自身、无法利用外生驱动变量的结构特性高度一致。

最后，从深度学习模型训练稳定性来看：

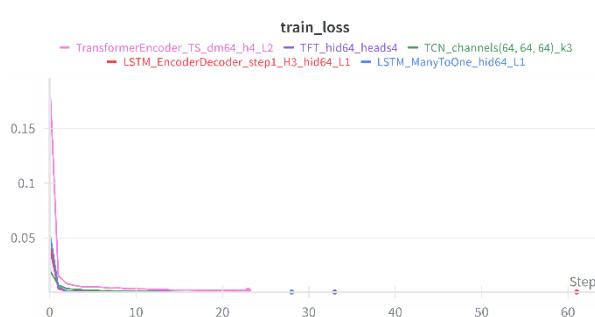


图34. train\_loss

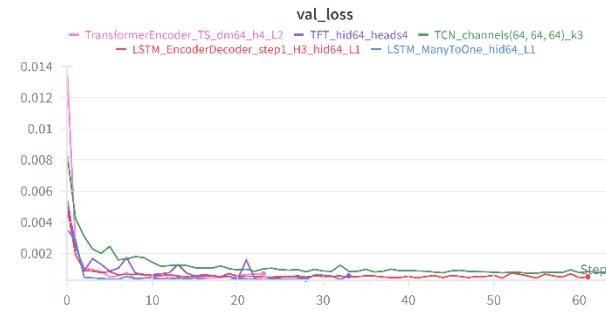


图35. val\_loss

大多数的深度学习模型的训练损失呈单调下降趋势，验证损失在训练中后期趋于稳定，没有出现明显的过拟合或震荡情况。LSTM 和 TFT 的验证损失曲线尤为平滑，说明在当前的网络规模和正则化强度下，模型已经达到了较为合适的复杂度。结合测试集评估结果，可以认为本研究中的主要深度学习模型在训练上是稳定的，预测性能具有一定的泛化能力。

## 7 结论

本研究基于 MODIS NDVI、CHIRPS 降水和 ERA5-Land 气象与土壤湿度等多源数据，构建了美国南部大平原干旱敏感草原区 2000–2024 年 8 日尺度的 NDVI–气象多变量时间序列数据集，在统一的数据预处理和滑动窗口框架下，系统化的比较了传统统计模型（Persistence、Climatology、线性回归、岭回归、ARIMA、SARIMA）与多种深度学习模型（LSTM 单步、LSTM Encoder–Decoder、TCN、Transformer Encoder、TFT）在 NDVI 预测及干旱相关低 NDVI 状态识别方面的性能。结果表明，多变量 LSTM 单步预测模型在

数值误差和干旱识别能力上都是表现最佳的模型，而 ARIMA/SARIMA 等单变量线性模型在本类型数据集上的表现是有明显的不足。

本实验的结果支持现有的关于 NDVI 对干旱具有时间尺度依赖性和滞后响应的认识，即植被对降水与土壤水分异常的响应会在一定时间滞后后才集中体现在 NDVI 的下降上。通过将降水、土壤湿度、气温和潜在蒸散与 NDVI 在统一时间尺度上进行了融合，并采用 12 个时间步的窗口作为输入，深度学习模型能够从多变量、长时间序列中自动提取非线性和时滞信息，相比仅仅依赖 NDVI 自身的 ARIMA/SARIMA 模型，更加适合于刻画干旱背景下植被动态的复杂过程。

在模型方面，多变量 LSTM 单步模型在 MAE、RMSE、F1-score 和 ROC-AUC 等指标上均处于最优或接近最优水平，说明该模型在数值拟合与低 NDVI 状态识别两方面都具有较强能力。岭回归和线性回归在多源特征输入下也表现出良好性能，表明在特征工程合理的情况下，简单线性模型仍然有重要的基线价值。TCN、Transformer Encoder 和 TFT 在误差与干旱识别方面都是明显优于单变量 ARIMA/SARIMA 的，并与线性模型处于同一性能梯队，显示出卷积和注意力结构在多源时间序列建模中的潜力。

## 8 总结

本研究的工作可以概括成三个层面。首先是在数据层面，基于 MODIS NDVI 产品构建了研究区长时间 NDVI 序列，并且将 CHIRPS 降水和 ERA5-Land 的土壤湿度、气温与潜在蒸散等日尺度变量按 8 日时间窗口聚合，统一到与 NDVI 一致的时间轴上。在时间划分上，采用时间顺序划分训练、验证与测试集，并且进行了缺失值插补与标准化处理；在特征工程上，通过引入年内日序的正余弦编码增强模型对季节周期的识别能力。

其次是在模型层面，在统一的 12 步滑动窗口框架下构建了多种模型体系，包括持久性模型、气候平均模型、多元线性回归与岭回归、ARIMA/SARIMA，以及 LSTM 单步、LSTM Encoder-Decoder、TCN、Transformer Encoder 和 TFT 等多种深度学习结构。所有模型都是在相同的训练 - 验证 - 测试划分和评估指标下进行的比较，从而保证了模型间对比的公平性。ARIMA/SARIMA 的引入，结合时间序列分析经典理论，为本研究提供了具有代表性的线性基准[9]；LSTM、TCN、Transformer 与 TFT 则代表了当前多源时间序列预测中的主流深度学习方法[5][6][7][8]。

最后在进行评估与分析时，通过 MAE 和 RMSE 两个回归误差指标以及基于 NDVI 低分位数阈值构建的 F1-score 和 ROC-AUC 两个分类指标，从数值拟合和干旱相关低 NDVI 状态识别这两个视角对本实验构建的模型进行了综合的评价。结合预测 - 真实对比图、残差时序图和训练 - 验证损失曲线，对模型的拟合能力、偏差结构和训练稳定性进行了系统分析。结果表明，多变量 LSTM 单步模型在整体性能上最为突出，线性回归与岭回归在多源特征条件下表现接近，TCN、Transformer Encoder 和 TFT 也具备一定的较强的预测能力，而单变量 ARIMA/SARIMA 表现是比较差的。

## 9 模型评价与优缺点

### 9.1 模型综合评价

综合上述模型的所有表现来看，多变量 LSTM 单步预测模型在本研究中表现最为均衡：在数值误差、干旱相关低 NDVI 状态识别和训练稳定性等方面都处于领先地位。该模型的优势主要源自长短期记忆结构能够在门控机制下有效保留和筛选与 NDVI 动态密切相关

的历史信息，既能刻画季节性和年际变率，又能对降水与土壤湿度的累积与滞后效应保持敏感[5]。

线性回归与岭回归的表现说明，在特征工程合理、特征维度适中且样本量有限的情况下，线性模型仍然可以具有较强的基线能力。它们能够利用降水、土壤湿度和温度等多源特征，在一定程度上刻画 NDVI 对气象条件的响应，这些模型的数值误差和干旱识别能力仅略逊于 LSTM 模型，是一个结构简单、易于解释的备选方案。

TCN、Transformer Encoder 与 TFT 在本研究中的表现也较为稳定，数值误差与干旱识别能力均明显优于持久性模型、气候平均模型和 ARIMA/SARIMA。TCN 通过膨胀卷积与残差网络在较少层数下实现了较长时间依赖的建模能力[6]；Transformer Encoder 和 TFT 则依赖注意力机制在时间步之间进行加权关联，能够在一定程度上识别对预测最关键的时间段和变量[8]。在当前数据规模和任务设置下，这些结构虽然整体上还是略逊于 LSTM 单步模型，但是已经展现出良好的泛化能力和可扩展性。

相对而言，ARIMA/SARIMA 虽然在刻画线性自相关和季节结构方面具有坚实的理论基础[9]，但是由于无法显式利用降水、土壤湿度等外生变量，且假设关系为线性，在本研究数据中难以捕捉 NDVI 对干旱的复杂响应。这种模型在干旱年份生长季期间的预测偏差是比较大的，F1-score 和 ROC-AUC 指标明显低于多源模型，因此更适合作为基线和对照，而非实际业务应用中的首选模型。

## 9.2 模型体系的优点与不足

本研究有一些主要的优点：

一是数据与模型框架统一，在同一滑动窗口、特征集合、评估指标和时间划分下比较多种模型，使得不同模型间的性能差异可以比较清晰且合理地归因于结构本身，而不是数据使用方式的不同。二是兼顾了传统统计模型和多种深度学习架构，从简单到复杂地构建了一个层次分明的模型谱系，便于在实际应用中根据数据规模、计算资源和解释需求选择适当复杂度的模型。三是引入了与干旱相关的低 NDVI 状态识别指标，使得模型评价不仅关注数值误差，也兼顾生态与风险监测需求，更加贴近干旱监测和生态管理的实际应用场景[10]。

本研究同样也存在一些不足：

首先，空间上采用区域平均 NDVI 和气象变量，忽略了像元尺度的空间异质性，未来可以在保持时间序列建模框架的同时，引入空间卷积或图神经网络刻画空间结构。其次，本文重点关注单步预测任务，多步预测和不确定性量化尚未系统展开，而多步预测对于季节尺度干旱预警和农业决策更具实际意义。再次，虽然引入了部分注意力结构（Transformer、TFT），但是没有充分挖掘这些模型的可解释性优势，未来可以进一步基于注意力权重分析不同时期和变量对干旱相关 NDVI 变化的相对贡献。

## 参考文献

- [1] Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived Normalized Difference Vegetation Index (NDVI) to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510.
- [2] Vicente-Serrano, S. M., Gouveia, C., Camarero, J. J., et al. (2013). Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences*, 110(1), 52–57.
- [3] Funk, C., Peterson, P., Landsfeld, M., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066.
- [4] Didan, K. (2021). MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500m. NASA LP DAAC / USGS.
- [5] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
- [6] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv preprint arXiv:1803.01271*.
- [7] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30.
- [8] Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. *International Journal of Forecasting*, 37(4), 1748–1764.
- [9] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
- [10] Vicente-Serrano, S. M., Quiring, S. M., Peña-Gallardo, M., et al. (2020). A review of environmental droughts: Increased risk under global warming? *Earth-Science Reviews*, 201, 102953.
- [11] Zhang, X., et al. (2024). The Response of NDVI to Drought at Different Temporal Scales in the Qinghai–Tibet Plateau. *Water*, 16(17), 2416.

# **Appendice**

## Appendix A GitHub代码仓库链接

[Stardust-12138/Final-Assignment---Time-Series-Modeling-of-Plant-Community-Responses-under-Drought-Conditions: 期末作业-干旱条件下植物群落响应的时间序列建模（231820002 陶德荣）Final Assignment - Time Series Modeling of Plant Community Responses under Drought Conditions](#)

## Appendix B Hugging Face数据集仓库链接

[TaoDerong/GreatPlains-Multisource-2000-2024 at main](#)

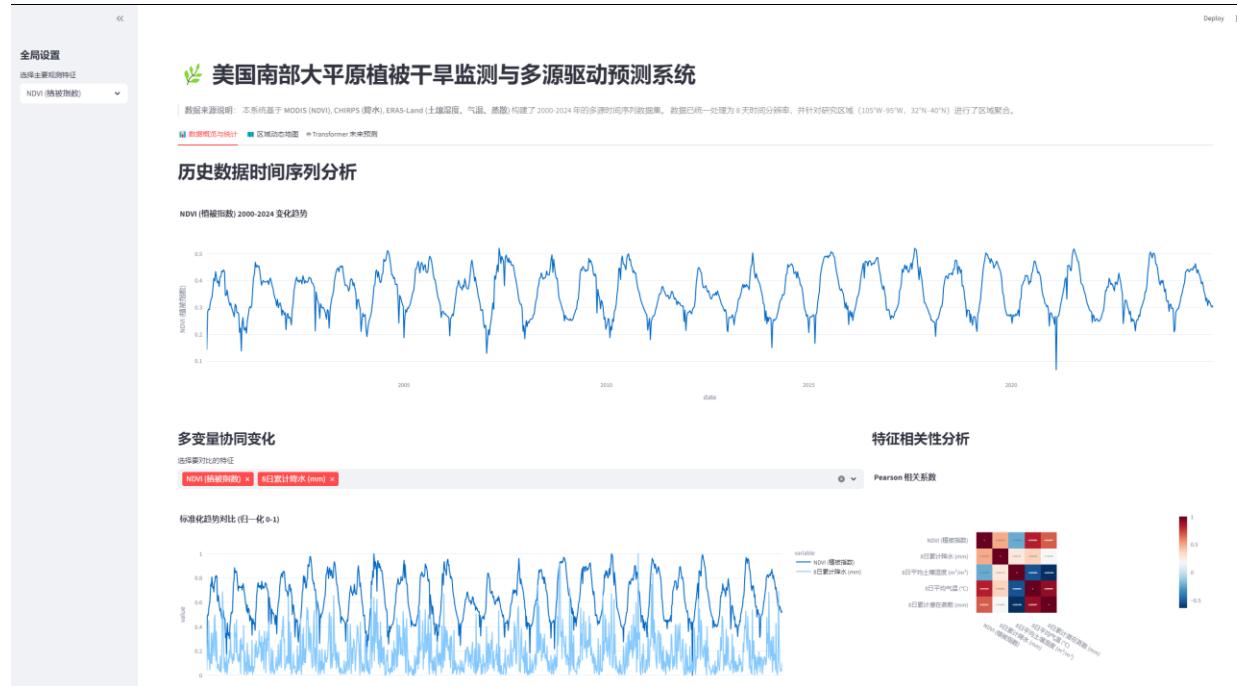
## Appendix C Wandb数据集仓库链接

[ndvi\\_drought\\_greatplains Workspace – Weights & Biases](#)

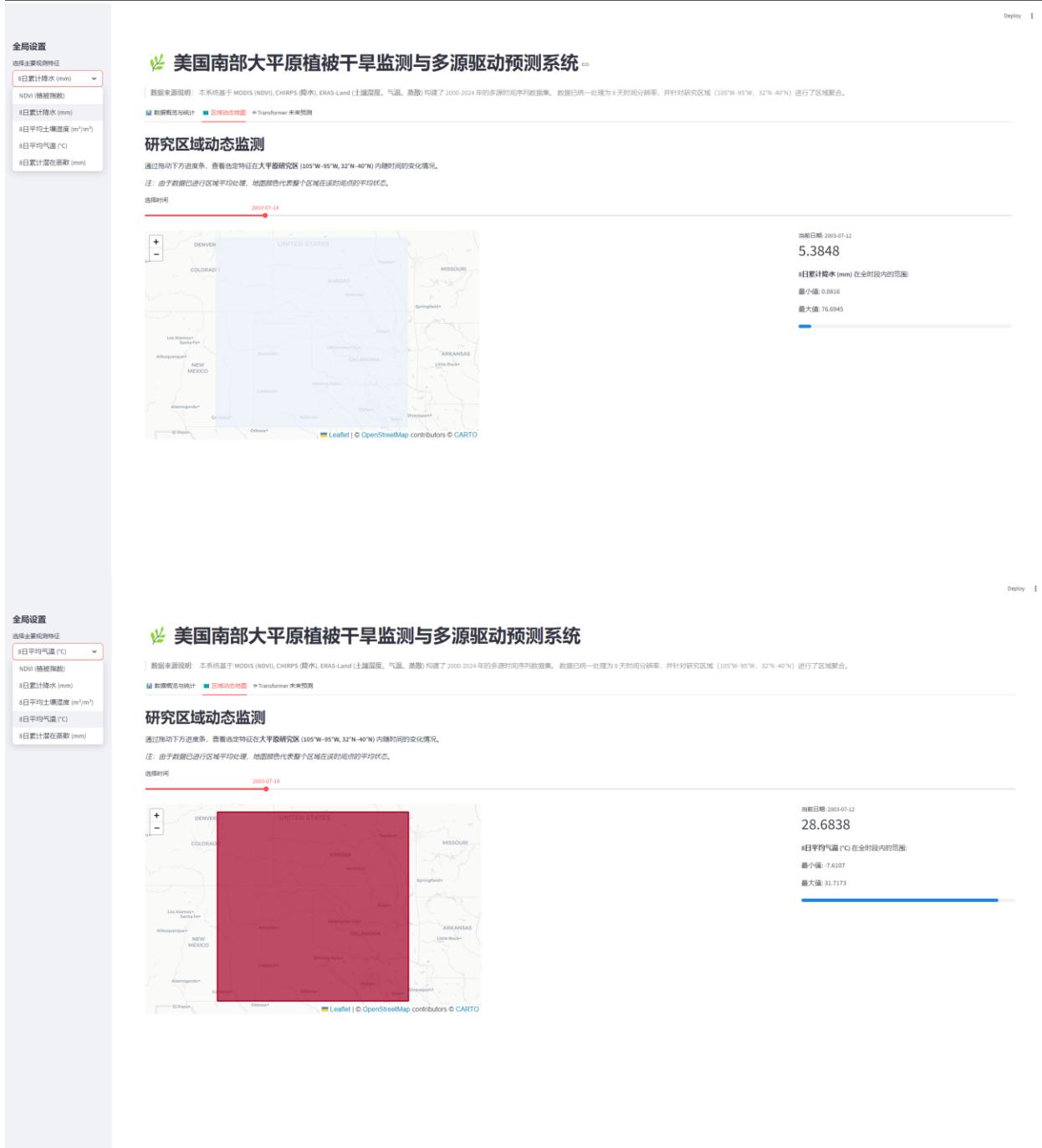
## Appendix D 可视化仪表盘

本次实验还构建了一个仪表盘，可在github仓库中找到app.py文件，预处理好的数据文件（csv文件）和对应的.bat文件（使用其他用户启动文件打开），一般只需要打开.bat文件根据提示安装对应库即可运行。

该仪表盘具体情况将截图放在下面进行参考：







**全局设置**

选择主要观测特征  
8日平均气温 (°C)

**多源驱动未来预测 (Transformer)**

本模块利用 Transformer 深度学习模型，学习 NDVI 与气象因子（降水、气温等）的历史时序关系。模型训练完成后，将执行多步滚动预测，生成超出当前数据集时间范围的未来趋势。  
训练完成！正在对未来推演...

**模型交互参数**

未来预测步数 (每步 8 天)  
12

回溯窗口大小 (Seq Length)  
13

训练轮次 (Epochs)  
50

Transformer 隐层深度  
64

**开始训练并预测**

**未来 96 天 8 日平均气温 (°C) 趋势预测**

Transformer 多步滚动预测结果: 8 日平均气温 (°C)

**查看模型训练损失 (Loss Curve)**

MSE Loss (每 Epoch) 下降情况，若曲线未收敛，请增加训练轮次。

---

**全局设置**

选择主要观测特征  
NDVI (植被指数)

**多源驱动未来预测 (Transformer)**

本模块利用 Transformer 深度学习模型，学习 NDVI 与气象因子（降水、气温等）的历史时序关系。模型训练完成后，将执行多步滚动预测，生成超出当前数据集时间范围的未来趋势。  
训练完成！正在对未来推演...

**模型交互参数**

未来预测步数 (每步 8 天)  
12

回溯窗口大小 (Seq Length)  
15

训练轮次 (Epochs)  
55

Transformer 隐层深度  
64

**开始训练并预测**

**未来 96 天 NDVI (植被指数) 趋势预测**

Transformer 多步滚动预测结果: NDVI (植被指数)

**查看模型训练损失 (Loss Curve)**

MSE Loss (每 Epoch) 下降情况，若曲线未收敛，请增加训练轮次。