

2021「AI Earth」

人工智能创新挑战赛 总决赛

神之一手YueTan



谭龙兴 岳宏颖

主办单位：南京信息工程大学 阿里巴巴达摩院 国家气候中心 国家海洋环境预报中心 安徽省气象局

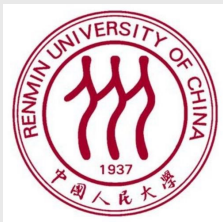
指导单位：中国气象服务协会

协办单位：国际气象教育与科学研究协会

团队成员



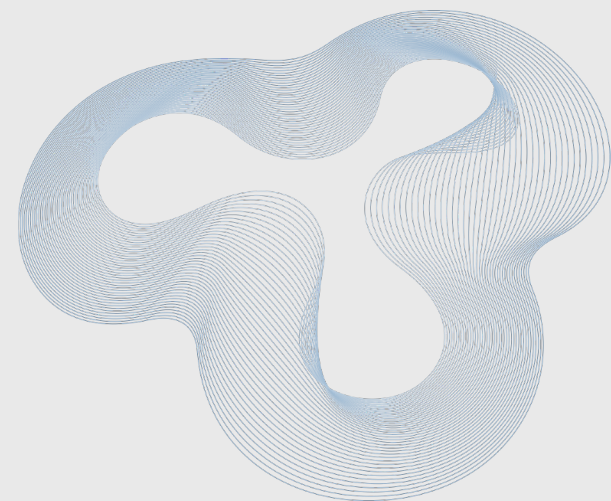
谭龙兴：浙江天垂，北京航空航天大学



岳宏颖：吉利汽车，中国人民大学



成绩：5th



任务分析

任务：

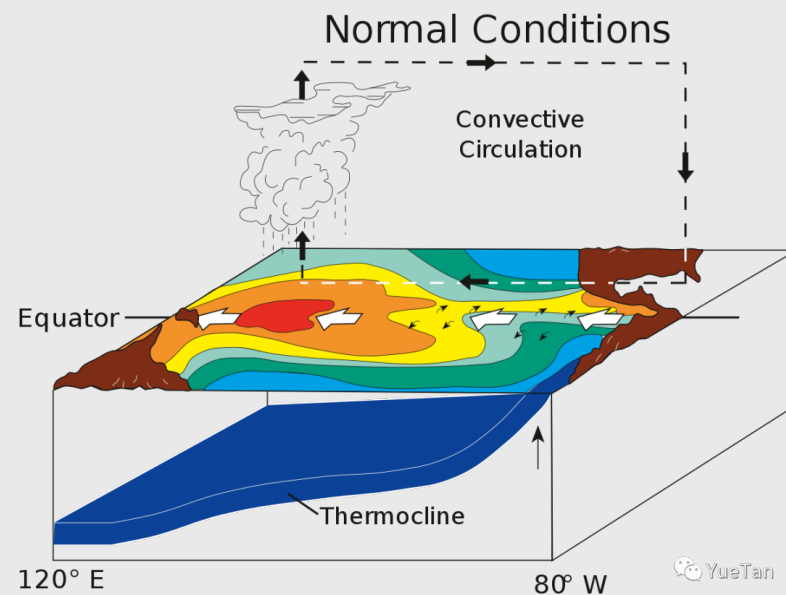
- 过去12个月的气象数据（sst、t300、ua、va）预测未来24个月的NINO3.4指标

评价函数：

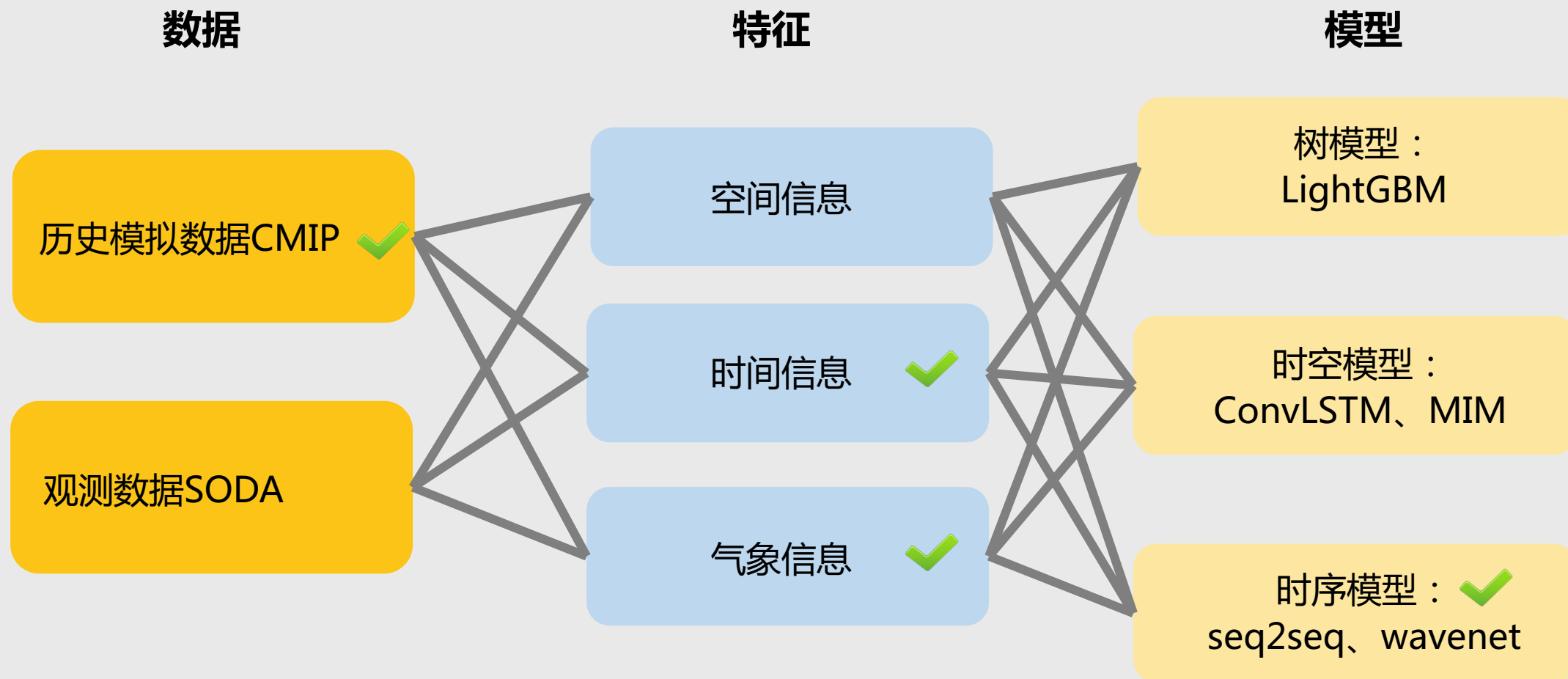
- 相关性表示整体趋势相似性，RMSE表示趋势的贴近程度，越远的预测权重越高

难点：

- 长距离预测
- 规律本身具有随机性（线上线下载分较大）



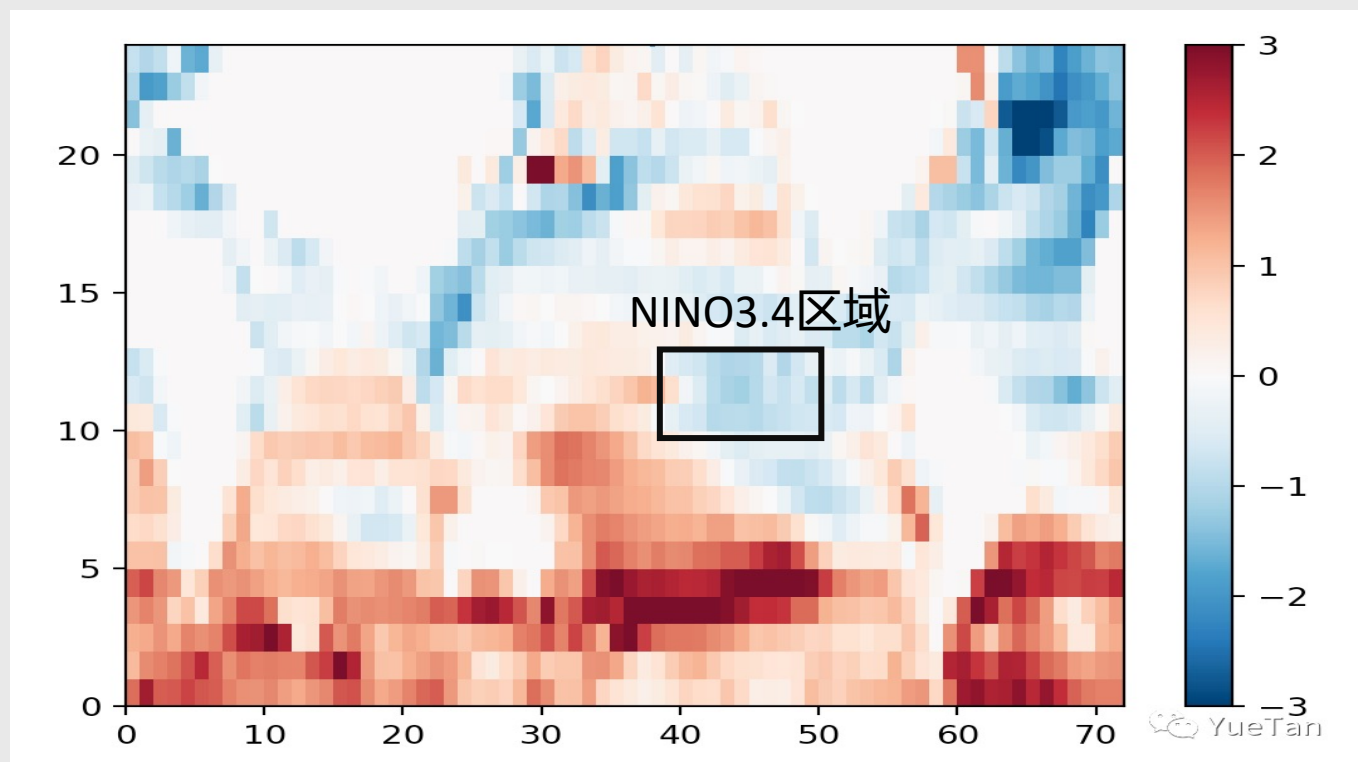
技术路线



最终方案：气象特征 + 简单模型

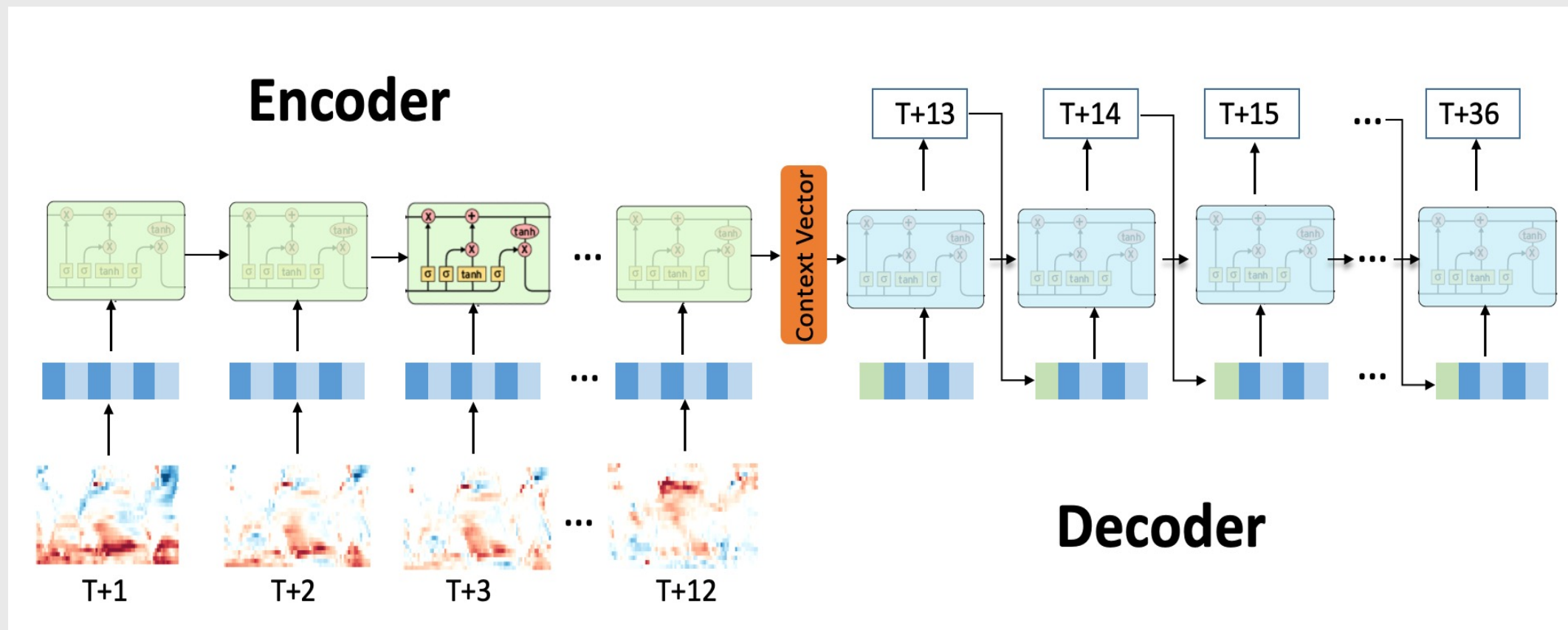
短期预测：全量数据+全局特征+时空模型

长期预测：采样数据+局部特征+时间序列模型



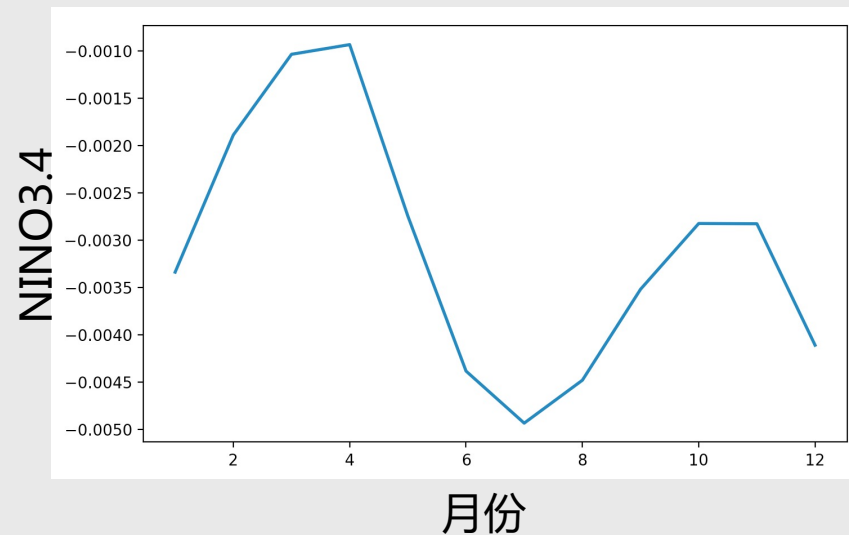
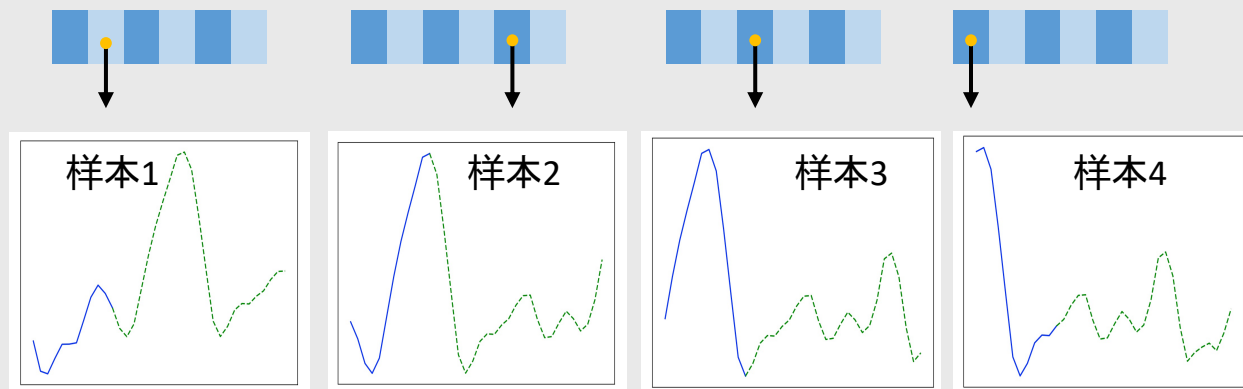
模型：简单高效

损失函数：根据评价函数增大长期预测权重， $\text{weight} = [\log 1p(i) \text{ for } i \text{ in range}(1, 25)]$



数据：CMIP训练，SODA验证

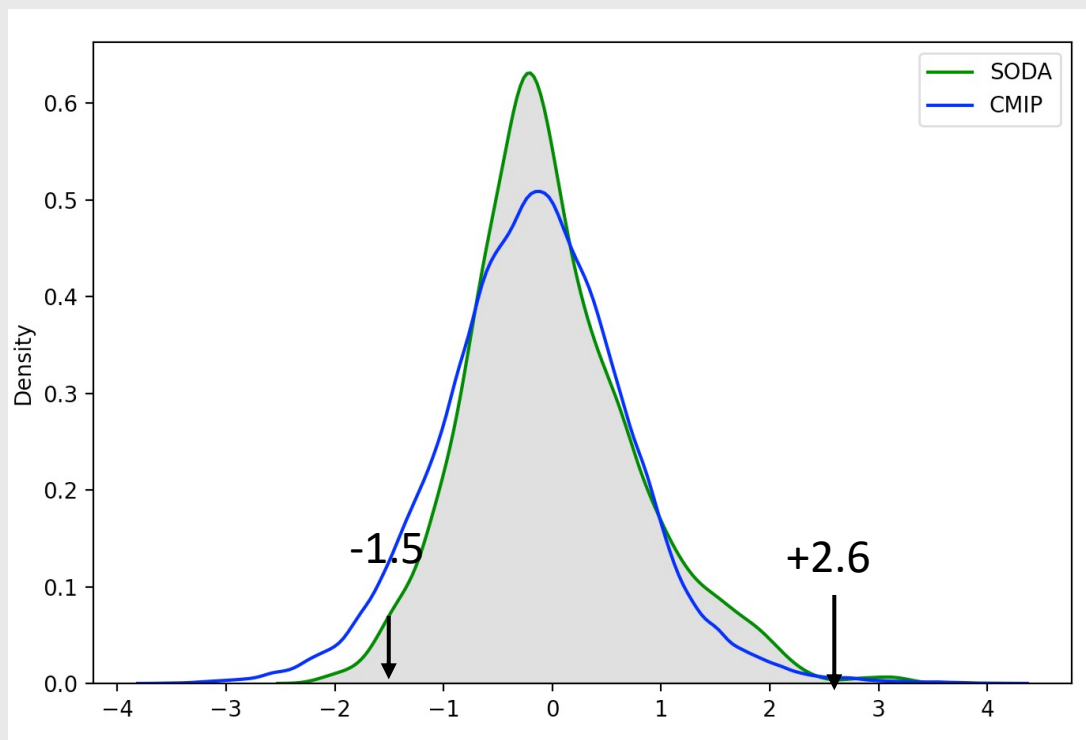
数据采样：每6个月随机选取一个起点作为样本



- NINO3.4大周期为3~4年、甚至7-8年，但每半年可能改变趋势具有小周期性
- 每6个月采样一次，每个时序样本中都蕴含新的变化规律，使样本具有代表性

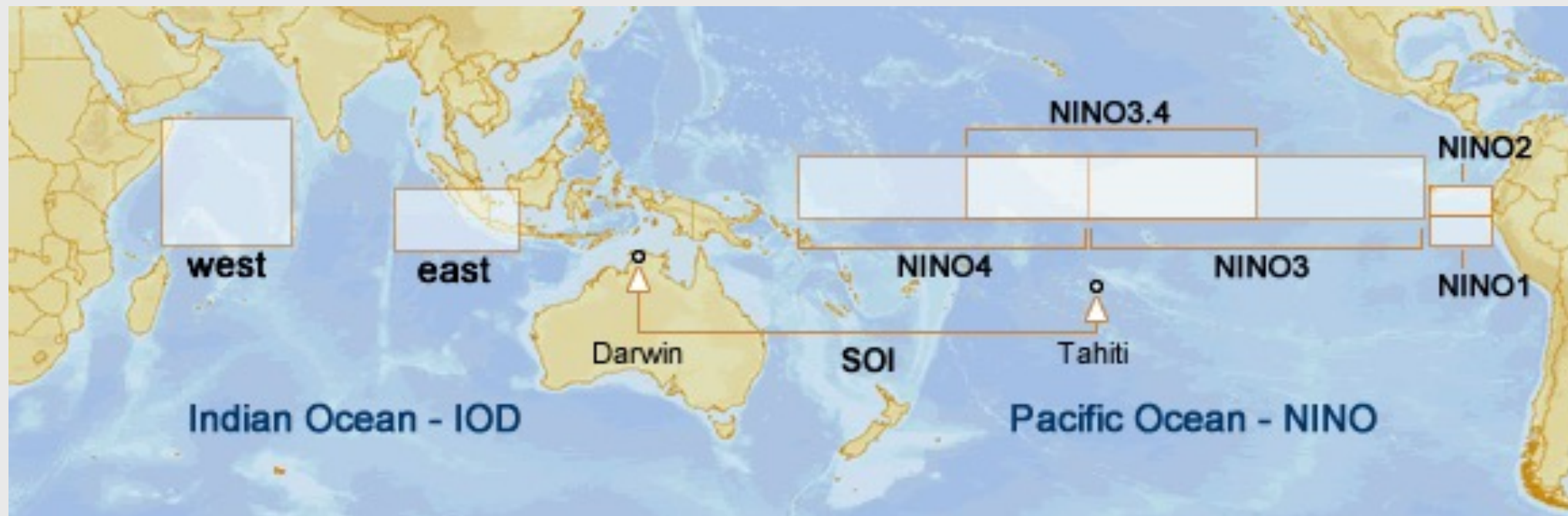
数据：CMIP训练，SODA验证

数据筛选：选择SODA近80年0.5% ~ 99.5%分位数，学习主体的变化规律



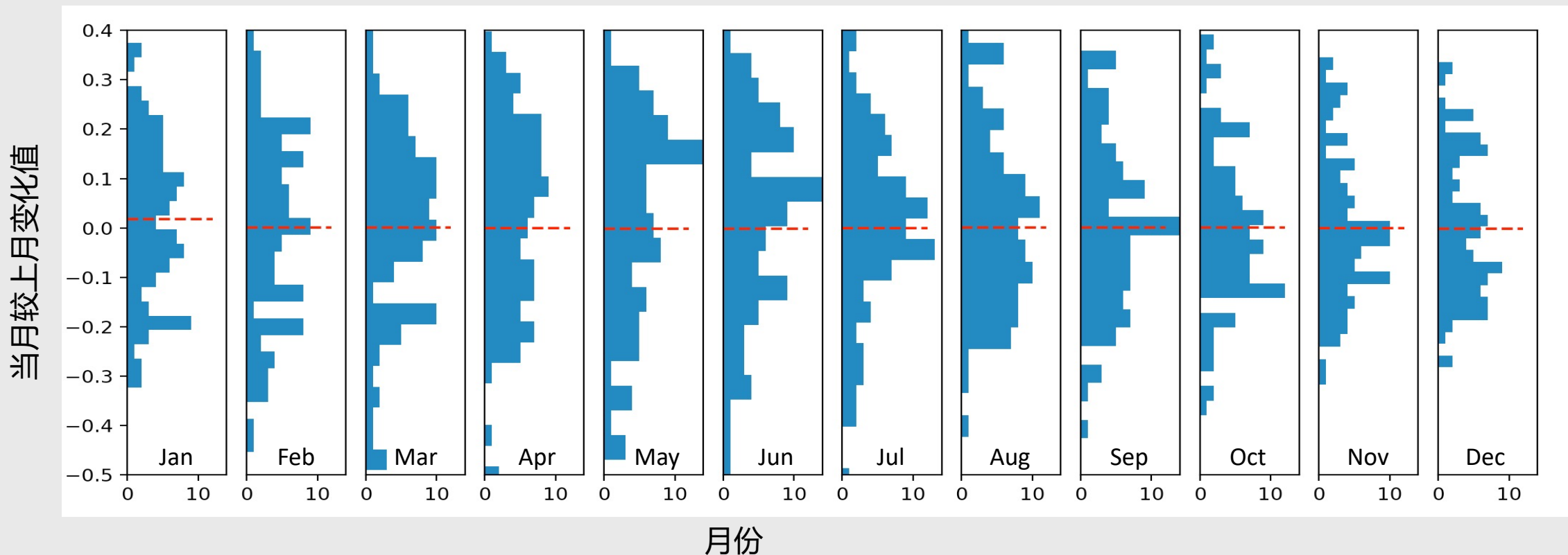
特征

气象特征: 善用领域知识

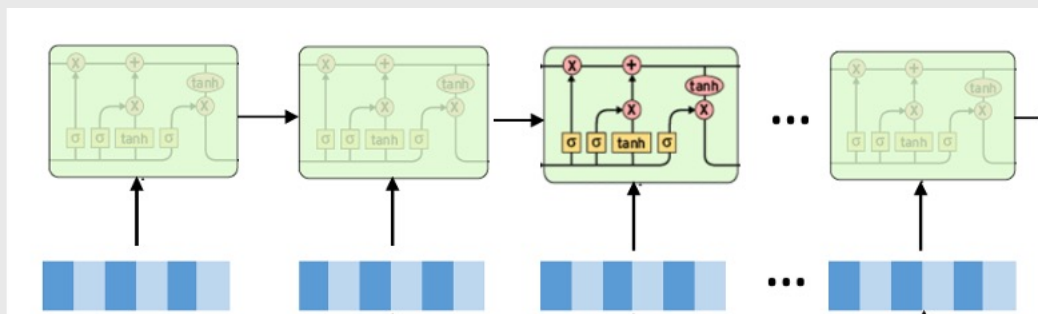


特征

时间信息特征: 不同月份与季度NINO3.4的变化情况不同 → 编码解码结构的模型



特征



编码特征：

气象特征：

NINO12 , NINO3 , NINO3.4 , NINO4 ,
TNI(NINO4-NINO12) , NINO4-NINO3

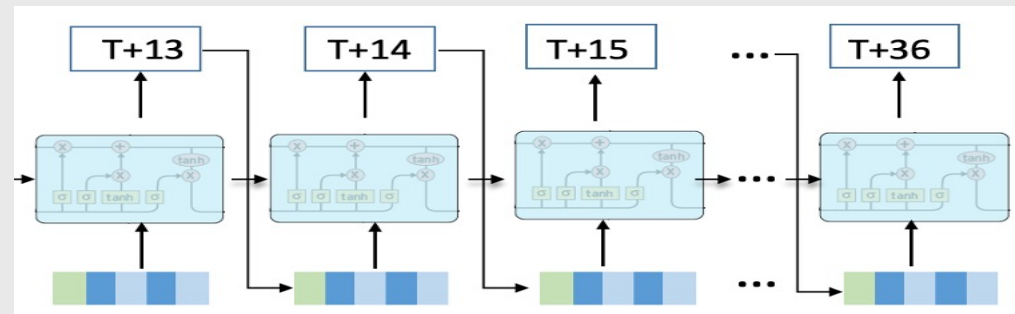
时序特征：

NINO3.4、T300前两个月历史值(lag2)

NINO3.4区域sst最大值、标准差

时间信息特征：

月份embedding、季度embedding



解码特征：

时间信息特征：

月份embedding、季度embedding

验证

验证方法：

- 对观测数据SODA进行十次采样，加最后一段的平均值

结果：

- 线上A榜与B榜分数相对一致
- 模型较大改动时，线上线下仍有不确定性
- 小改动时，线下结果可以较好反应线上成绩

目标傅立叶变换：

- 对目标傅立叶变换可以提取目标间的波形关系，同时将时域中24个值的预测转化为频域中5个值的预测

时间预训练方法：

- 取前36个月历史预测未来24个月作为预训练权重，训练时将前24Dropout为零，留下后12个

数据和特征增强：

- 倒序，mixup增强，信号处理中的特征提取

总结

模型简单高效：

气象特征加简单编码解码模型，线上训练加预测半小时内结束

善用领域知识：

注重气象知识融入到特征，增加可解释性

深入数据特点：

根据数据特点，决定数据的采样规律，使样本具有代表性

The background is a deep blue with a complex, abstract pattern of overlapping geometric shapes, primarily squares and rectangles, creating a sense of depth and movement. Thin, white, curved lines sweep across the composition, adding a dynamic feel. The overall aesthetic is modern and technological.

谢谢
THANKS