

# 地表温度时空融合模型性能对比：地表环境、空间尺度差异及数据相关性的影响分析

陈欣远, 余楚楚, 王猛猛, 张正加, 李冉, 李阳阳, 刘修国

中国地质大学(武汉) 地理与信息工程学院, 武汉 430078

**摘要:** 时空融合是当前结合影像高空间分辨率和高时间分辨率两种属性最有效的方法, 对精细化遥感数据产品的生产具有重要意义。然而, 绝大多数时空融合模型都是基于地表反射率和植被指数开发, 较少应用于地表温度 LST (Land Surface Temperature); 另外, 现有的地表温度时空融合模型评价未充分考虑不同数据组合与数据相关性对模型性能的影响。为此, 本文基于 4 种类型的时空融合模型 (增强型自适应反射融合模型 ESTARFM (Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model)、时空反射解混模型 STRUM (Spatial and Temporal Reflectance Unmixing Model)、灵活时空数据融合 FSDAF (Flexible Spatiotemporal Data Fusion) 模型和增强型深度卷积时空融合网络 EDCSTFN (Enhanced Deep Convolutional Spatiotemporal Fusion Network), 比较各模型的地表温度融合结果, 同时分析地表环境、数据空间尺度差异以及数据相关性对融合结果的影响。结果表明: (1) EDCSTFN 模型融合 LST 精度最高, 其次是 FSDAF 和 ESTARFM 模型, STRUM 模型精度相对较差; ESTARFM 和 EDCSTFN 模型的融合影像目视效果更优, STRUM 和 FSDAF 模型的融合影像相对较为平滑; (2) 4 种模型在地物结构单一区域融合效果最理想, 其次是物候变化明显区域, 在地表覆盖类型复杂区域融合效果欠佳。EDCSTFN 模型在不同地表环境下均保持高水平精度, 尤其是地表覆盖类型复杂区域, 但在地物结构简单的区域表现不如 ESTARFM 模型; (3) 随着空间尺度差异的增大, ESTARFM、STRUM 和 FSDAF 模型的融合精度表现为规律性下降, 而 EDCSTFN 模型并没有下降的趋势; (4) 随着数据相关性的降低, 4 种模型的融合精度均呈现规律性降低, 其中 EDCSTFN 模型表现出相对较好的稳定性和鲁棒性。

**关键词:** 时空融合, 地表温度, 增强型自适应反射融合模型, 时空反射解混模型, 灵活时空数据融合模型, 增强型深度卷积时空融合网络

**中图分类号:** TP701/P2

**引用格式:** 陈欣远, 余楚楚, 王猛猛, 张正加, 李冉, 李阳阳, 刘修国. 2025. 地表温度时空融合模型性能对比: 地表环境、空间尺度差异及数据相关性的影响分析. 遥感学报, 29(12): 3575–3594

Chen X Y, She C C, Wang M M, Zhang Z J, Li R, Li Y Y and Liu X G. 2025. Performance comparison of spatiotemporal fusion models for land surface temperature: Analysis of the effect of surface environment, spatial scale differences, and data correlation. National Remote Sensing Bulletin, 29(12): 3575–3594 [DOI: 10.11834/jrs.20254400]

## 1 引言

地表温度 LST (Land Surface Temperature) 作为了解陆地—大气系统地表能量与水分平衡的重要方式 (Duan 等, 2020; Sobrino 等, 2016), 在蒸散发估算 (Li 等, 2017; Chen 和 Liu, 2020)、土壤水分反演 (Long 等, 2019; Song 等, 2022)、气候变化分析 (Peng 等, 2014; Zhou 等, 2018)、城

市热环境分布 (Weng, 2009; Xian 等, 2022) 以及能量平衡建模 (Bateni 和 Entekhabi, 2012; Silva 等, 2018) 等领域得到了广泛应用。热红外遥感技术因其准确性高、限制少以及探测范围大等优势成为了获取地表温度产品的最佳方式 (Mao 等, 2017); 同时, 随着遥感技术的不断进步, 可以及时、有效地获取海量高空间分辨率与高时间分辨率地表温度数据; 然而, 传感器物理限制、数据

收稿日期: 2024-09-26; 预印本: 2025-08-29

基金项目: 国家自然科学基金 (编号: 61801443, 42571374); 国家重点研发计划 (编号: 2016YFA0600302); 湖北省自然科学基金 (编号: 2025AFB691)

第一作者简介: 陈欣远, 研究方向为遥感数据重建与时空融合。E-mail: cxy15390627@cug.edu.cn

通信作者简介: 王猛猛, 研究方向为定量遥感反演和遥感信息智能提取。E-mail: wangmm@cug.edu.cn

处理和存储需求以及信号噪声比等的影响使单一传感器获取的影像无法同时具备高空间分辨率与高时间分辨率两种属性 (Chen 等, 2023)。例如, Landsat 8—9 卫星热红外波段的空间分辨率为 100 m, 但重访周期却长达 16 d; 而 Terra/Aqua 卫星组合重访周期仅为 0.5 d, 但搭载的中分辨率成像光谱仪 MODIS (Moderate-resolution Imaging Spectroradiometer) 传感器的热红外波段空间分辨率却高达 1000 m。时空融合技术是一种用于提升遥感数据时空分辨率的方法, 其利用已知时间点不同时空分辨率的数据在时间上的变化特征和空间上的尺度差异, 来预测特定时间点的高时空分辨率数据 (Wu 等, 2021)。在不改变现有观测条件的前提下, 时空融合技术可以有效结合两种类型影像的优势, 提供更加精确、全面的信息 (Xiao 等, 2023), 因此在精准作物产量估计 (Meng 等, 2018, 2019)、精细化土地覆盖分类 (Chen 等, 2017; Yin 等, 2012)、生物量精确测算 (Zhang 等, 2015; Zhou 等, 2023)、地表蒸散发连续估计 (Wang 等, 2018b)、植被生长精准监测 (Gao 等, 2017, 2015; Xie 等, 2018; Zhang 等, 2021) 以及云去除 (Li 等, 2019) 等领域发挥了重要作用。

根据技术原理与侧重点的不同, 可将时空融合模型分为权重模型、解混模型、贝叶斯模型、学习模型和混合模型 5 类 (Zhu 等, 2018)。权重模型通过权重函数赋予邻域像素权重并求和来估算目标像素的值。Gao 等 (2006) 率先提出的时空自适应反射融合模型 STARFM (Spatial and Temporal Adaptive Reflectance Fusion Model) 通过移动窗口搜索相似像元, 并从时间、空间和光谱差异 3 个角度衡量权重并对窗口中心像元值进行预测。Zhu 等 (2010) 提出的增强型时空自适应反射融合模型 ESTARFM (Enhanced STARFM) 通过引入转换系数来对 STARFM 模型进行改进, 提高了对于异质区域的预测精度。Cheng 等 (2017) 提出的时空非局部滤波融合模型 STNLFFM (Spatial and Temporal Nonlocal Filter-based Fusion Model) 通过引入非局部滤波提高了时空冗余信息的利用率以及预测精度。Wang 等 (2018a) 提出的 Fit-FC 模型通过回归模型拟合 (RM)、空间滤波 (SF) 以及残差补偿 (RC) 3 个步骤提高了在地物剧烈变化区域的预测效果。解混模型基于线性光谱混合理论对低分辨率像素进行降阶。Wu 等 (2012) 提出的时空

数据融合方法 STDFA (Spatial Temporal Data Fusion Approach) 通过在一个移动窗口内对参考和预测日期的端元反射率进行解混来估计反射率变化, 然后将估计的变化添加到参考高分辨率图像中获得预测。Guo 等 (2015) 提出了一种改进的基于解混的融合方法 IUBF (Improved Unmixing-Based Fusion), 在传统解混方法的基础上, 通过波段选择、滑动窗口分类和引入插值图像等改进, 生成更适合内陆水体遥感监测的融合图像。贝叶斯模型通过贝叶斯最大后验概率估计 (MAP) 使预测影像最接近真实影像。Huang 等 (2013) 在提出的一种光谱融合和时空融合相统一的贝叶斯框架 (Unified fusion) 中, 采用低通滤波模拟空间差异, 用线性函数表示时间变化, 在已知影像的条件下, 使用贝叶斯估计理论预测影像的条件概率密度函数, 通过 MAP 估计未知参数, 从而计算预测影像。学习模型通过字典或机器学习方法对高低分辨率影像间的非线性关系进行建模。Huang 和 Song (2012) 提出的基于稀疏表示的时空地表反射融合模型 SPSTFM (Sparse-representation-based Spatio-temporal Reflectance Fusion Model) 通过字典学习的方式融合参考影像并对结果进行预测。Tan 等 (2019) 基于深度学习卷积网络提出了增强型深度卷积时空融合网络 EDCSTFN (Enhanced Deep Convolutional Spatiotemporal Fusion Network) 在融合高低分辨率影像的过程中表现出较好的灵活性和鲁棒性, 融合精度优于多数经典算法。混合模型指采用包括权重函数、像元解混思想、贝叶斯理论和学习模式在内的两种或两种以上构造策略的时空融合方案。Gevaert 和 García-Haro (2015) 基于混合像元分解的思想, 并采用贝叶斯理论来约束解混过程, 提出了时空反射解混模型 STRUM (Spatial and Temporal Reflectance Unmixing Model)。Zhu 等 (2016) 在 STARFM 模型的基础上, 通过结合混合像元分解与薄板样条插值提出了灵活时空数据融合 FSDAF (Flexible Spatiotemporal Data Fusion) 模型。

尽管已有模型均被证明具有独特的优势, 但数据来源、区域特征以及参数设置等的不同使研究者难以确定哪种方法性能最优。鉴于此, Liu 等 (2019) 使用 Landsat 7 ETM+ 与 MODIS 地表反射率数据分析了景观异质性和时间变化对 STARFM、UBDF、One-Pair Learning Method、FSDAF 和 Fit-

FC 5 种模型性能的影响, 结果显示 FSDAF 模型最稳健, Fit-FC 模型效率最高, One-Pair Learning Method 模型适用于地表覆盖发生明显变化的区域。Zhou 等 (2021) 使用归一化植被指数 NDVI (Normalized Difference Vegetation Index) 时间序列数据了几何配准、辐射差异和空间分辨率 3 个因素对 UBDF、LMGM、STARFM、Fit-FC、OPDL 和 FSDAF 6 种模型性能的影响, 结果显示 Fit-FC 最适用于 NDVI 预测。雷晨阳等 (2021) 使用地表反射率、植被指数和地表温度数据集, 从单时相、时间序列和多数数据产品角度对 STARFM、ESTARFM、FSDAF 和 STNLFFM 4 种模型进行综合评价, 结果显示 STARFM 不适用于地物复杂的情况, STNLFFM 融合效果最好。

然而, 与地表反射率和 NDVI 数据相比, 地表温度数据具有时序变化剧烈、波动性强、地类敏感性较低、光谱特征匮乏以及像元相关性较强等特点 (李冉等, 2022), 因此其时空预测更具挑战性。考虑到不同数据融合原理的相似性, 已有多基于地表反射率和 NDVI 开发的时空融合模型在地表温度数据中得到了应用, 如 STARFM (Li 等, 2017), ESTARFM (Ma 等, 2018) 和 FSDAF (杨敏等, 2018) 等。在此基础上, 许多学者根据地表温度的特点也进行了新的探索。Weng 等 (2014) 和 Wu 等 (2015) 对 STARFM 模型进行了改进, 分别提出了用于温度制图的时空自适应数据融合算法 SADPAT (Spatio-temporal Adaptive Data Fusion Algorithm for Temperature Mapping) 和时空综合温度融合模型 STITFM (Spatio-temporal Integrated Temperature Fusion Model), 有效解决了太阳辐射年振荡和两种不同轨道卫星产品空间尺度差异对融合结果的影响。为了获取逐小时空间连续的地表温度数据, Wang 等 (2024) 和 Ding 等 (2023) 将随机森林模型应用于时空融合, 同时引入再分析数据, 弥补了云下 LST 的数据缺失。Tang 等 (2023) 将空间填充与时空融合相结合, 进一步提高了地表温度重建的精度。此外, 也有部分研究针对不同时空融合模型应用于地表温度数据的表现进行了评估, Li 等 (2020) 使用极轨和对地静止卫星地表温度数据对 STARM、基于解混的融合方法 UBF (Unmixing-Based Fusion)、FSDAF 和 STITFM 4 种模型进行比较, 结果显示 FSDAF 模型性能最佳。Tang 等 (2024) 采用 STARFM、基于

空间加权的虚拟图像对融合 VIPSTF-SW (Spatial Weighting-based Virtual Image Pair Fusion) 和 FSDAF 3 种模型来融合地表温度数据, 结果显示 STARFM 更适合用于生成 100 m 的逐小时地表温度数据, 同时基于单图像对的时空融合模型能够避免在融合过程中引入额外的不确定性。总体而言, 传统模型中权重模型和混合模型往往在地表温度时空融合任务中表现更为突出, 学习模型则在解决数据缺失问题中展现出较大的应用潜力。然而, 大多数现有对比研究中所使用的模型类型并不全面, 且在不同场景下的适用性仍有待进一步探索。此外, 目前对时空融合模型性能的评价绝大多数都基于一种高低分辨率组合的参考影像对且通常选择参考影像对时仅考虑参考时刻和预测时刻之间的时间跨度, 并未充分考虑不同空间尺度组合与参考时刻和预测时刻间的数据相关性对模型性能的影响。

因此, 本文选取 4 种不同类型的经典时空融合模型 (ESTARFM、STRUM、FSDAF 和 EDCSTFN), 分别对应权重模型、解混和贝叶斯混合模型、解混和权重混合模型以及学习模型, 4 种模型涵盖了 5 类时空融合模型的核心思想, 可以更加全面地比较不同类型的时空融合模型应用于地表温度数据融合时的差异。本研究基于 Landsat 8—9 和 MODIS 数据建立长时序地表温度数据集, 通过地面实测数据和实际 Landsat LST 影像数据对 4 种模型的融合精度进行验证与分析, 同时分别在不同地表环境、不同空间尺度数据组合以及不同数据相关性的条件下进行对比实验, 评估 4 种模型在不同场景下的准确性与适用性, 进而为后续应用模型选取与融合模型优化提供参考。

## 2 研究区域与数据

### 2.1 研究区域

为比较 4 种时空融合模型在不同地表环境下的地表温度融合结果, 本研究共选取地物结构单一、地表覆盖类型复杂以及地表物候变化明显的 3 个研究区域。其中, A 区域主要位于青藏高原 ( $32^{\circ}39'57''\text{N}$ — $33^{\circ}04'48''\text{N}$ ,  $82^{\circ}40'34''\text{E}$ — $83^{\circ}10'01''\text{E}$ ), 地表覆盖类型为草地和裸地, 包含极少数的林地, 地物结构较为单一。B 区域主要位于甘肃省张掖市 ( $38^{\circ}41'56''\text{N}$ — $39^{\circ}06'43''\text{N}$ ,  $100^{\circ}12'58''\text{E}$ — $100^{\circ}44'42''\text{E}$ ),

区域内不透水地表、水体、耕地、草地和裸地等高度混合,地表覆盖类型较复杂。C区域主要位于北京市大兴区(39°34′50″N—39°59′21″N, 116°15′47″E—116°47′31″E),区域内地表覆盖类型包括不透水地

表和耕地,具有明显的物候变化。3个研究区域的地表覆盖类型以及B区域内两个气象观测站大满超级站(CJZ)和机场荒漠站(JCHM)的空间分布如图1所示。

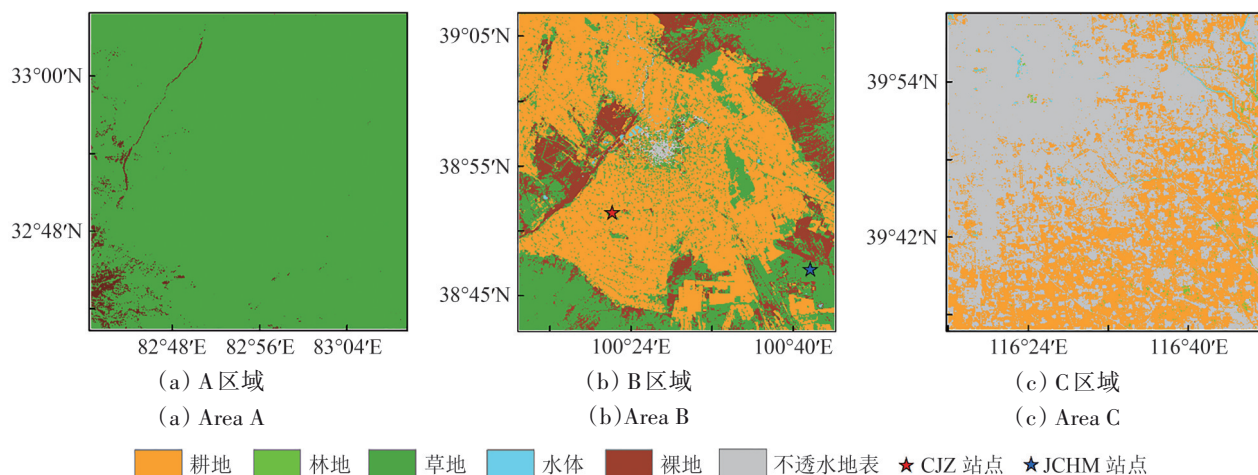


图1 3个研究区域地表覆盖类型及地面站点位置分布

Fig. 1 Land cover types and location distribution of ground stations in the three study areas

## 2.2 实验数据

### 2.2.1 地表温度数据集

本文构建的地表温度数据集由长时序高低分辨率影像对组成。数据集使用的高分辨率影像数据为Landsat 8—9的L2级别数据,其中的热红外波段地表温度产品经过了大气校正处理,空间分辨率为30 m,投影方式采用UTM-WGS84投影。低分辨率影像数据来自自由Terra卫星搭载的MODIS传感器获取的每日地表温度产品MOD11A1,空间分辨率为1000 m,投影方式采用Sinusoidal正弦曲线投影。Landsat数据在3个研究区的轨道号分别为143/037、133/033和123/032(列/行),MODIS的轨道号分别为24/05、25/05和26/05(列/行),所有实验数据均通过USGS Earth Explorer (<https://earthexplorer.usgs.gov/>[2024-09-26])网站公开下载获得。

为确保Landsat LST数据和MODIS LST数据在空间上能够匹配,首先将Landsat数据分别裁剪至3个研究区范围,影像大小统一为1504×1504。MODIS数据重投影为UTM-WGS84投影后使用双线性内插法进行重采样,再将其裁剪至与Landsat数据相同范围,得到一一对应的Landsat-MODIS

LST影像对。其中,A区域共收集2013年—2023年41对LST影像对,B区域共收集2014年—2021年44对LST影像对,C区域共收集2017年—2023年40对LST影像对,所有影像数据均无云覆盖,质量良好。图2为构建的地表温度数据集示例,图2中展示了3个研究区部分日期的Landsat-MODIS LST影像对,同一日期的Landsat LST影像和MODIS LST影像为一组,每个研究区分别展示7组LST影像对。

### 2.2.2 地面实测数据

本研究使用的地面实测数据来自“黑河流域生态水文过程综合遥感观测联合试验HiWATER(Heihe watershed allied telemetry experimental resaerch)”。HiWATER是一项综合性生态水文实验,旨在提高水文和生态过程的可观测性(Liu等,2018)。HiWATER地面站点主要测量3类变量,包括关键生态水文过程的状态变量和通量、大气强迫数据和其他参数(如植被、土壤、地形、水文和空气动力参数),站点测量结果已用于多种LST产品验证(Li等,2023;Wang等,2023)。本次验证采用了两个HiWATER站点,分别为CJZ(38.8555°N, 100.3722°E,下垫面为耕地)和JCHM

(38.7781°N, 100.6967°E, 下垫面为荒漠)。其中CJZ 站点获取了2014年2月—2021年5月期间的测量数据, JCHM站点获取了2014年2月—2018年10月期间的测量数据, 两个站点均位于B区域内, 具体位置分布如图1所示。

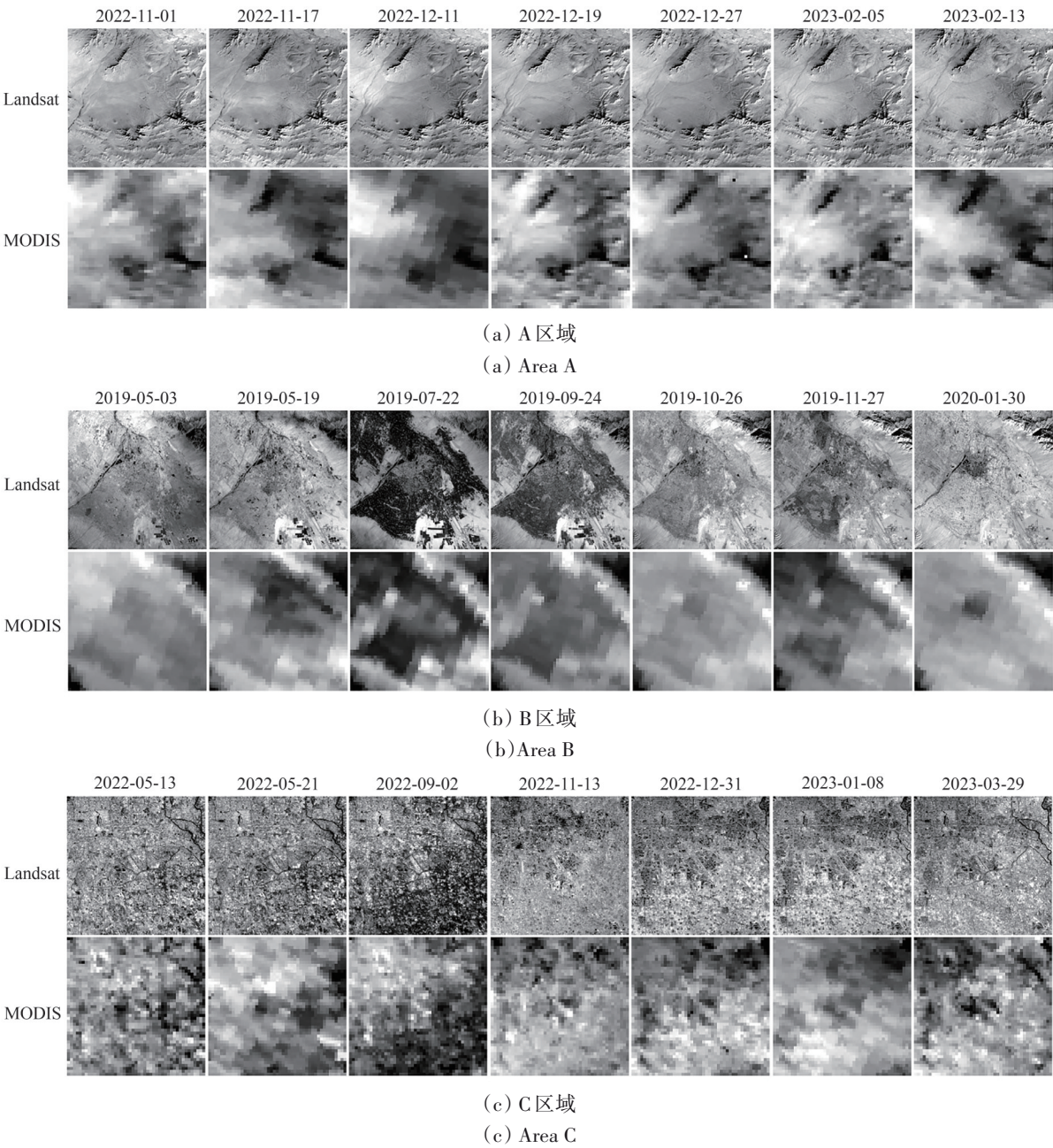


图2 不同研究区 Landsat-MODIS 地表温度数据集示例  
Fig. 2 Example of Landsat-MODIS LST data sets of different study areas

3 研究方法

3.1 时空融合方法

3.1.1 ESTARFM模型

ESTARFM模型 (Zhu等, 2010) 针对STARFM模型 (Gao等, 2006) 在复杂地表下预测精度较差

的问题进行了改进。ESTARFM模型认为纯净的高分辨率影像像元间存在回归关系, 即:

$$F_{x_i, y_j, t_k} = a \times C_{x_i, y_j, t_k} + b \tag{1}$$

式中,  $F_{x_i, y_j, t_k}$  和  $C_{x_i, y_j, t_k}$  分别表示高低分辨率影像在  $t_k$  时刻  $(x_i, y_j)$  像元位置处的像元值,  $a$  和  $b$  代表回归系数。

ESTARFM模型假设一段时间内影像像元值线

性变化、地物丰度保持不变且同类低分辨率像元转换系数相同,通过两对已知时相 $t_0$ 和 $t_2$ 的高低分辨率影像作为参考影像对和一景预测时刻 $t_1$ 的低分辨率影像共同作为输入影像,以移动窗口为单位对整幅影像进行逐像元融合。在权重计算时,ESTARFM模型综合考虑了空间差异与光谱差异,同时结合时间权重对两期预测结果进行融合,以提高预测结果的准确性。

$$F_{x_{w/2}, y_{w/2}, t_p} = F_{x_{w/2}, y_{w/2}, t_k} + \sum_{i=1}^N W_i \times V_i \times (C_{x_i, y_i, t_p} - C_{x_i, y_i, t_k}) \quad (2)$$

$$F_{x_{w/2}, y_{w/2}, t_p} = \sum T_k \times F_{x_{w/2}, y_{w/2}, t_k}, k = m, n \quad (3)$$

式中,  $F_{x_{w/2}, y_{w/2}, t_p}$  代表预测时刻 $t_p$ 最终的高分辨率影像像元预测结果,  $F_{x_{w/2}, y_{w/2}, t_k}$  代表参考时刻 $t_k$ 单期的预测结果,  $N$ 代表包括中心像元在内的相似像元个数,  $w$ 为滑动窗口大小,  $(x_{w/2}, y_{w/2})$ 代表待测像元中心位置,  $(x_i, y_i)$ 代表相似像元的坐标,  $W_i$ 为第 $i$ 个相似像元的权重,  $V_i$ 代表当前窗口中心像元的转换系数,  $T_k$ 代表参考时刻 $t_k$ 的时间权重。

### 3.1.2 STRUM模型

STRUM模型 (Gevaert 和 García-Haro, 2015) 是一种基于混合像元分解思想的融合算法,并在混合像元分解过程中使用贝叶斯定理约束光谱分解可能出现的不准确的地物类别光谱。STRUM模型需要一对已知时相 $t_0$ 的高低分辨率影像和一景预测时刻 $t_1$ 的低分辨率影像作为参考影像。主要步骤如下:

(1) 对 $t_0$ 时刻的高分辨率影像进行 ISODATA 非监督分类。

(2) 对 (1) 的分类结果应用移动窗口识别窗口内的地物类别数以及丰度信息,丰度小于阈值的地物合并到光谱相似的类别中。

(3) 计算低分辨率影像在移动窗口内从 $t_0$ 到 $t_1$ 的像元变化值 $\Delta C$ ,并进行混合像元分解,采用贝叶斯理论约束解混过程,以尽可能减少光谱解混中因不确定的光谱类别所造成的残差 $\varepsilon$ 。

(4) 将通过解混得到的每一个地物类别从 $t_0$ 到 $t_1$ 的变化值 $\Delta F_k$ 分配到 $t_0$ 时刻的高分辨率影像中得到 $t_1$ 时刻的预测结果。

$$\Delta C = \sum A_k \times \Delta F_k + \varepsilon \quad (4)$$

$$F_{t_1} = F_{t_0} + \Delta F \quad (5)$$

式中,  $A_k$ 代表第 $k$ 个地物类别的丰度值。

### 3.1.3 FSDAF模型

FSDAF模型 (Zhu 等, 2016) 通过借鉴像元解混和加权函数的思想提高了在地表覆盖渐变和突变地区的预测精度。FSDAF模型需要的参考影像与STRUM模型一致。主要步骤如下:

(1) 对 $t_0$ 时刻的高空间分辨率影像采用 ISODATA 分类法确定低分辨率像元中每个地物类型的丰度信息。

(2) 根据光谱线性分解理论,选择在 $t_0$ 到 $t_1$ 时间段内无地表覆盖类型变化的 $n$ 个像元用以确定每个类的时间变化 $\Delta F_k$ 。

(3) 将 $t_0$ 时刻高分辨率影像像元值加上时间变化得到 $t_1$ 时刻的时间预测结果,同时引入残差 $\varepsilon_{TP}$ 进行修正来提高预测精度。

(4) 进行时间预测后采用薄板样条TPS (Thin Plate Spline) 插值方法将 $t_1$ 时刻的低空间分辨率像元降尺度到高空间分辨率来进行空间预测,得到空间尺度上的残差 $\varepsilon_{SP}$ 。

(5) 通过设置权重函数,将最终残差合理地分配给步骤 (2) 中的时间变化,从而提高时间预测的精度。

(6) 使用阈值法确定相似像元后,结合空间差异计算中心像元的总变化并与 $t_0$ 时刻高分辨率影像像元相加得到每个像元最终的预测值。

$$F_{t_1} = F_{t_0} + \sum_{c=1}^n W_c \times (\Delta F + \varepsilon) \quad (6)$$

式中,  $W_c$ 代表相似像元的权重,  $c$ 代表相似像元个数,  $\varepsilon$ 代表加权后的残差值。

### 3.1.4 EDCSTFN模型

EDCSTFN模型 (Tan 等, 2019) 是一种基于深度卷积网络的端对端的时空融合模型,通过从高一低空间分辨率的数据中自动提取特征信息,同时在图像特征层级进行信息融合,最后将融合特征还原到像素层级实现时空融合。

EDCSTFN的整体架构由3个子网络构成,分别为高空间分辨率影像编码器 (LTHS Encoder)、残差编码器 (Residual Encoder) 和重建解码器 (Reconstruction Decoder)。3个子网采用相同的网络结构,均由3个卷积层构成,且特征图在所有层中保持其原始空间分辨率。首先,通过高空间分辨率影像编码器提取参考日期高空间分辨率影像 $F_{t_1}$

的特征; 其次, 基于参考日期的高低空间分辨率影像对  $F_{t_0}$  和  $C_{t_0}$  以及预测日期的低空间分辨率影像  $C_{t_1}$ , 通过残差编码器提取参考日期与预测日期之间的高空间分辨率下的特征变化信息; 最后, 将参考日期的高空间分辨率特征图和高空间分辨率下的特征变化信息特征图进行相加得到预测日期的影像特征图, 通过重建解码器将预测日期的特征图映射到原始的像素空间, 得到预测日期的高空间分辨率影像。该过程可表示为

$$F_{t_1} = D(E(F_{t_0}) + R(F_{t_0}, C_{t_0}, C_{t_1})) \quad (7)$$

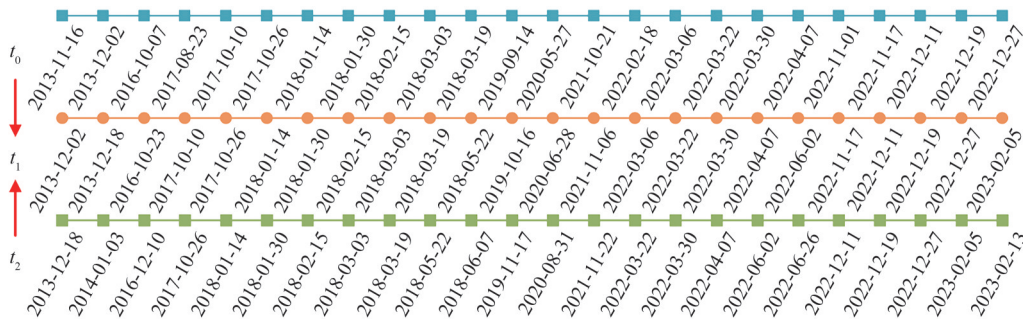
式中,  $E()$  代表高空间分辨率影像编码器,  $R()$  代表残差编码器,  $D()$  代表重建解码器。

模型的损失函数采用一种综合内容损失、特征损失和视觉损失的复合损失函数, 综合考虑融合精度损失和图像质量损失。其中, 内容损失通过计算模型预测结果和观测真值的均方误差 MSE (Mean Squared Error) 进行评估; 特征损失通过引入预训练的自编码器, 将模型预测结果和观测真值分别输入到自编码器中, 计算两者输出结果的

MSE 来评估; 视觉损失使用多尺度结构相似性 MS SSIM (Multi Scale Structural Similarity) 来评估融合结果和观测真值的相似性。

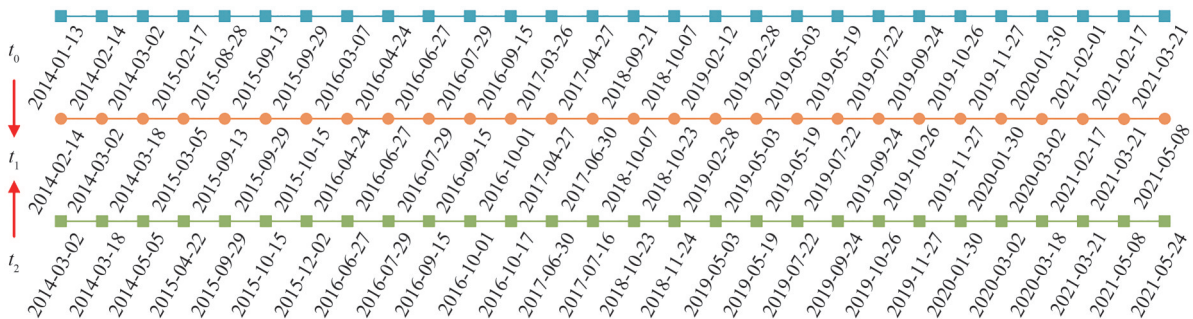
### 3.2 融合对比方案

时空融合模型的输入包括参考时刻的 Landsat-MODIS LST 影像对和预测时刻的 MODIS LST 影像, 一般选择与预测时刻相邻的日期作为参考时刻, 这种情况下, 参考影像和预测影像之间的温度变化、地表覆盖类型变化、传感器差异变化较小, 能够有效提高预测精度 (Li 等, 2021)。将预处理得到的 Landsat-MODIS LST 影像对按时间顺序排序, 并根据时间距离分组, 每组包括  $t_0$ 、 $t_1$  和  $t_2$  3 个时间点的影像对, 其中  $t_0$  和  $t_2$  为参考时刻,  $t_1$  为预测时刻, 分组结果如图 3 所示。由于 ESTARFM 和 EDCSTFN 模型在进行融合实验时均需要两对参考影像, 因此同一组的两对参考影像均作为模型输入; 而 STRUM 和 FSDAF 仅需要一对参考影像, 因此在  $t_0$  和  $t_2$  中选择与  $t_1$  更接近的时间点作为参考影像输入模型。



(a) A 区域

(a) Area A



(b) B 区域

(b) Area B

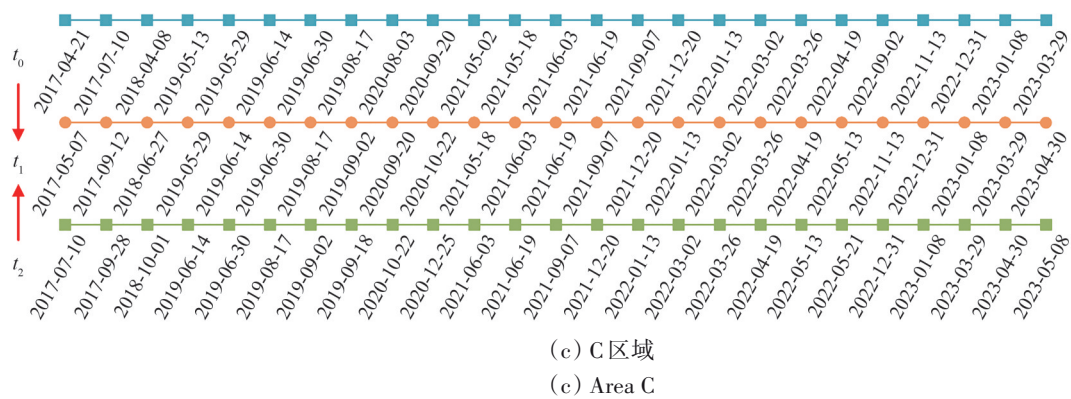


图3 3个研究区参考时刻和预测时刻分组结果  
Fig. 3 Grouping results of reference time and prediction time in three study areas

为比较4种时空融合模型在不同空间尺度数据组合条件下的地表温度融合结果，本研究需使用多种高低分辨率数据组合进行实验。为保证数据空间尺度差异的连续性，本研究选取30 m、120 m、240 m、480 m和960 m空间分辨率的数据进行组合，分别通过平均聚合的方法得到120 m、240 m和480 m的Landsat LST数据，通过双线性内插法重采样得到960 m、480 m、240 m和120 m的MODIS LST数据。将处理得到的高低分辨率LST数据进行（高一低）组合得到30—120 m、30—240 m、30—480 m、30—960 m、120—960 m、240—960 m和480—960 m共7种空间分辨率组合的Landsat-MODIS LST影像对。在比较不同地表环境对4种模型融合结果的影响时，为了避免较大的空间尺度变换对实验结果产生额外影响，本研究选取与原始LST产品空间分辨率最为接近的30—960 m组合的LST影像对作为实验数据，开展3个研究区域所有预测时刻的融合实验，在比较不同空间尺度数据

组合对4种模型融合结果的影响时，选择A区域的2022-12-19、B区域的2019-09-24和C区域的2022-11-13共3个时间点作为测试日期进行融合实验。

考虑到地表温度的时序波动性，时间距离难以较好地描述不同影像之间的相似性，本研究利用数据相关性代替时间跨度，比较4种时空融合模型在不同数据相关性条件下的地表温度融合结果。实验选取30—960 m空间分辨率组合的LST影像对，分别统计3个测试日期前后各3个时间点的MODIS LST与测试日期MODIS LST之间的相关系数（ $R$ ），根据相关系数大小重新将影像对进行分组，分组结果如表1所示。其中 $R_{01}$ 和 $R_{21}$ 表示参考时刻 $t_0$ 和参考时刻 $t_2$ 的MODIS LST分别与预测时刻 $t_1$ 的MODIS LST的相关系数。对于使用一对参考影像的STRUM和FSDAF模型选择相关系数更高的LST影像对作为模型输入。

表1 3个测试日期参考时刻和预测时刻再分组结果  
Table 1 Regrouping results of reference time and prediction time for the three test dates

| 区域 | 预测时刻 $t_1$ | 分组序号 | 参考时刻 $t_0$ | 相关系数 $R_{01}$ | 参考时刻 $t_2$ | 相关系数 $R_{21}$ |
|----|------------|------|------------|---------------|------------|---------------|
| A  | 2022-12-19 | 1    | 2022-11-01 | 0.7599        | 2023-02-05 | 0.8149        |
|    |            | 2    | 2022-12-11 | 0.7560        | 2022-12-27 | 0.7837        |
|    |            | 3    | 2022-11-17 | 0.6653        | 2023-02-13 | 0.6814        |
| B  | 2019-09-24 | 1    | 2019-07-22 | 0.7806        | 2019-10-26 | 0.8274        |
|    |            | 2    | 2019-05-19 | 0.5931        | 2019-11-27 | 0.6962        |
|    |            | 3    | 2019-05-03 | 0.4699        | 2020-01-30 | 0.6046        |
| C  | 2022-11-13 | 1    | 2022-09-02 | 0.6726        | 2022-12-31 | 0.7489        |
|    |            | 2    | 2022-05-21 | 0.3509        | 2023-01-08 | 0.5955        |
|    |            | 3    | 2022-05-13 | 0.1202        | 2023-04-30 | 0.0307        |

### 3.3 参数设置

本研究中，4种时空融合模型的参数设置参考已有研究（Zhu等，2010，2016；Gevaert和García-Haro，2015；Tan等，2019；雷晨阳等，2021）并根据实际数据进行调整和选择，为了有效比较4种模型的性能，对于功能相似的部分参数设置为相同值。ESTARFM模型移动窗口大小为51，地物类别数为4；FSDAF模型移动窗口大小为51，相似像元数为20，最小、最大类别分别为4和6；STRUM模型移动窗口大小为51，最小、最大类别分别为4和6，类别丰度阈值为0.05；EDCSTFN模型需要将数据集划分为训练集和测试集，为了评价所有预测时刻的融合结果，采用K折交叉验证方法。根据3个区域数据集的数量（24、28和25），将训练集和测试集的比例分别设置为5:1、3:1和4:1，因此3个区域的K分别设置为6、4、5。此外，超参数设置Adam优化算法的初始学习率为0.001，批次大小为4，每个实验组进行60次训练。

### 3.4 评价指标

融合质量评价是时空融合的后续，并可为模型选择、参数优化等提供必要的支持。本研究基于地面实测数据和实际Landsat LST影像数据对4种时空融合模型的融合结果进行综合评价，采用决定系数（ $R^2$ ）、均方根误差RMSE（Root Mean Square Error）、峰值信噪比PSNR（Peak Signal-to-Noise Ratio）和结构相似度（Structural Similarity, SSIM）4个统计指标进行定量评价。

$R^2$ 用于评估预测值与真实值的拟合程度，其取值范围为 $[0, 1]$ ， $R^2$ 的值越高，表示两组像素之间的一致性越强。RMSE可以衡量变量间的差异并直观反映离散程度，同时对变量间的异常值较为敏感，其取值范围为 $[0, +\infty)$ ，RMSE的值越小，表示融合精度越高。PSNR是基于误差敏感的图像质量评价，其与人眼的视觉特性不同，因此评价结果往往与人的主观感觉不一致。本文中PSNR主要用来衡量时空融合结果的噪声水平，PSNR的值越大，表示融合图像质量越高。其计算公式为

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X_{ij} - Y_{ij})^2 \quad (8)$$

$$PSNR = 10 \times \lg \frac{MAX^2}{MSE} \quad (9)$$

式中， $X_{i,j}$ 和 $Y_{i,j}$ 分别代表 $(i, j)$ 像元位置处的真实像素值和预测像素值，MSE代表均方误差， $H$ 、 $W$

分别为影像的高度和宽度，MAX代表影像中的最大像素值。

SSIM是一种视觉评价指标，可以从亮度、对比度和结构3个方面衡量两幅影像的相似度。SSIM的取值范围为 $[0, 1]$ ，值越接近1代表融合影像与实际影像的相似度越高。其计算公式为

$$SSIM = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (10)$$

式中， $\mu_X$ 、 $\mu_Y$ 、 $\sigma_X$ 、 $\sigma_Y$ 和 $\sigma_{X,Y}$ 分别代表两幅影像的均值、方差以及协方差， $C_1$ 和 $C_2$ 为两个常数，避免计算结果的不稳定。

## 4 结果与分析

### 4.1 模型验证及对比

本节采用B区域30—960 m空间分辨率组合的LST影像对并根据图3的分组进行融合实验，以地面站点实测数据作为真值对融合结果进行验证，结果见图4。可见4种模型的融合LST与地面实测LST之间都具有较高的相关性，决定系数 $R^2$ 均高于0.9。其中：EDCSTFN的融合精度最高， $R^2$ 达到了0.962，RMSE为2.358 K；其次是FSDAF和ESTARFM模型，两种模型均显示出较好的融合精度且差距极小， $R^2$ 和RMSE分别为0.927/0.922和3.741 K/3.763 K；STRUM模型表现相对欠佳，虽然 $R^2$ 达到了0.907，但是RMSE为4.117 K，误差相对较大。

为了更详细地比较4种模型的融合性能，图5展示了4种模型在预测时刻2019-09-24的融合结果影像与实际Landsat LST影像。从整体影像可以看出：ESTARFM和EDCSTFN模型的融合影像在视角效果上更接近实际影像，能够较好地保留地表温度分布的空间结构信息；STRUM和FSDAF模型的融合结果相对较差，大部分区域的地表温度预测值明显低于实际影像。从局部放大影像可以看出：ESTARFM和EDCSTFN模型都较好地保留了影像中的地物边界细节，但对于部分地表温度较低的区域，ESTARFM模型的预测值存在高估的现象，而EDCSTFN模型在温度变化较为剧烈的区域，存在部分细节丢失；STRUM的融合结果相对较为平滑，温度梯度保留较差，影像显得更加均匀；FSDAF模型的表现与STRUM相似，影像的细节部分也较为平滑，但在部分区域（如温度变化较小的区域）相比于STRUM模型能够保留更多的细节。

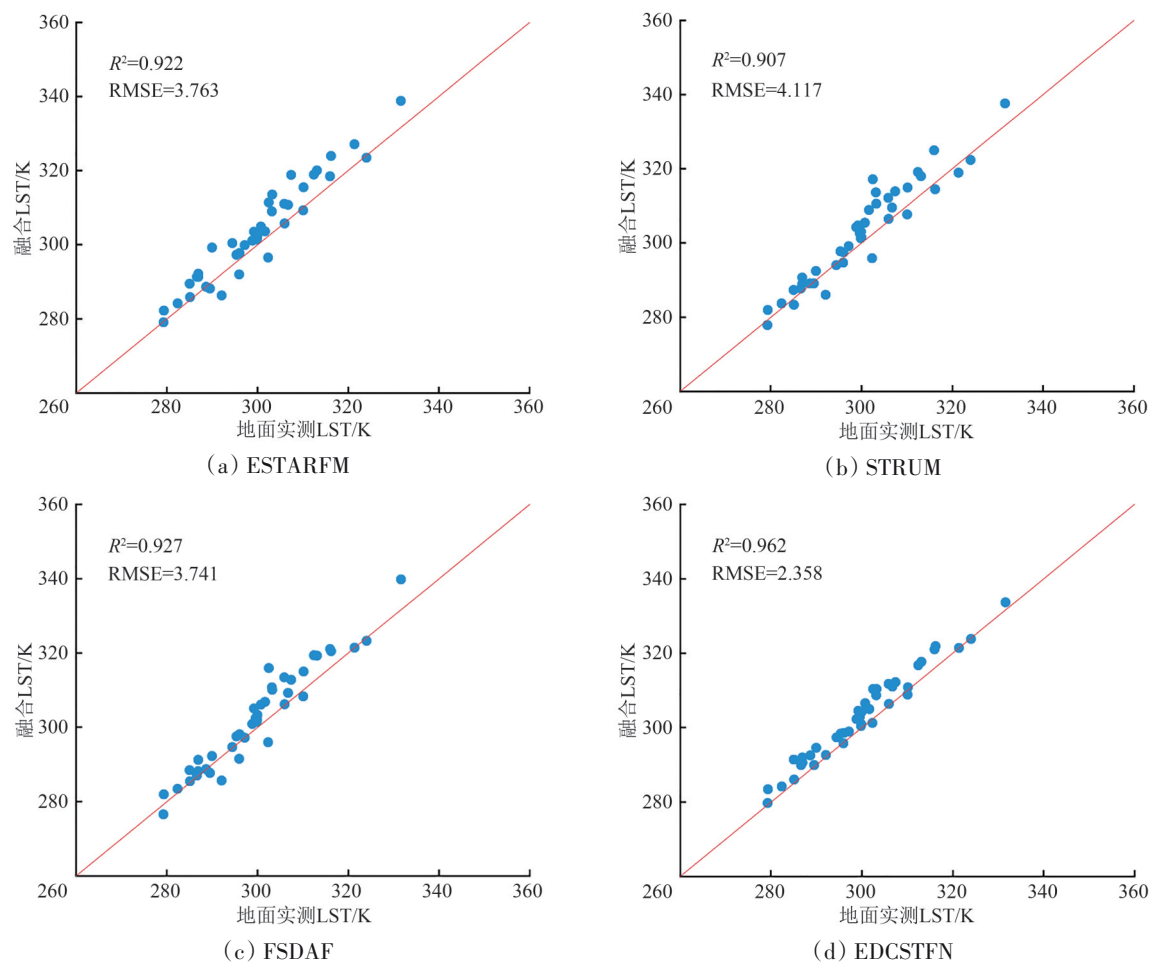


图4 不同时空融合模型下融合LST和地面实测LST的散点图

Fig. 4 Scatter plot of fused LST and ground measured LST under different spatiotemporal fusion models

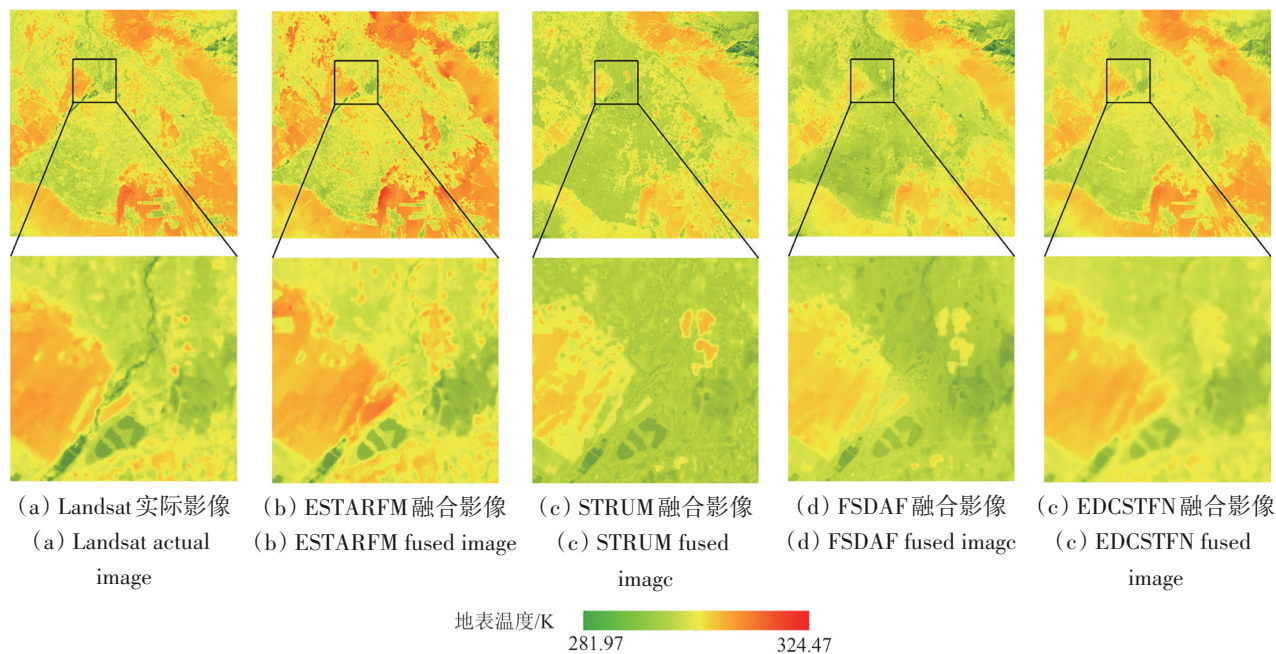


图5 4种时空融合模型的融合LST影像和实际Landsat LST影像

Fig. 5 Fused LST images of four spatiotemporal fusion models and actual Landsat LST image

## 4.2 地表环境对融合结果的影响

本节采用3个研究区30—960 m空间分辨率组合的LST影像对并根据图3的分组进行融合实验,

以实际 Landsat LST 影像作为真值对融合结果进行精度评价。图6为4种模型分别在3个研究区所有预测时刻的融合结果的平均量化结果。

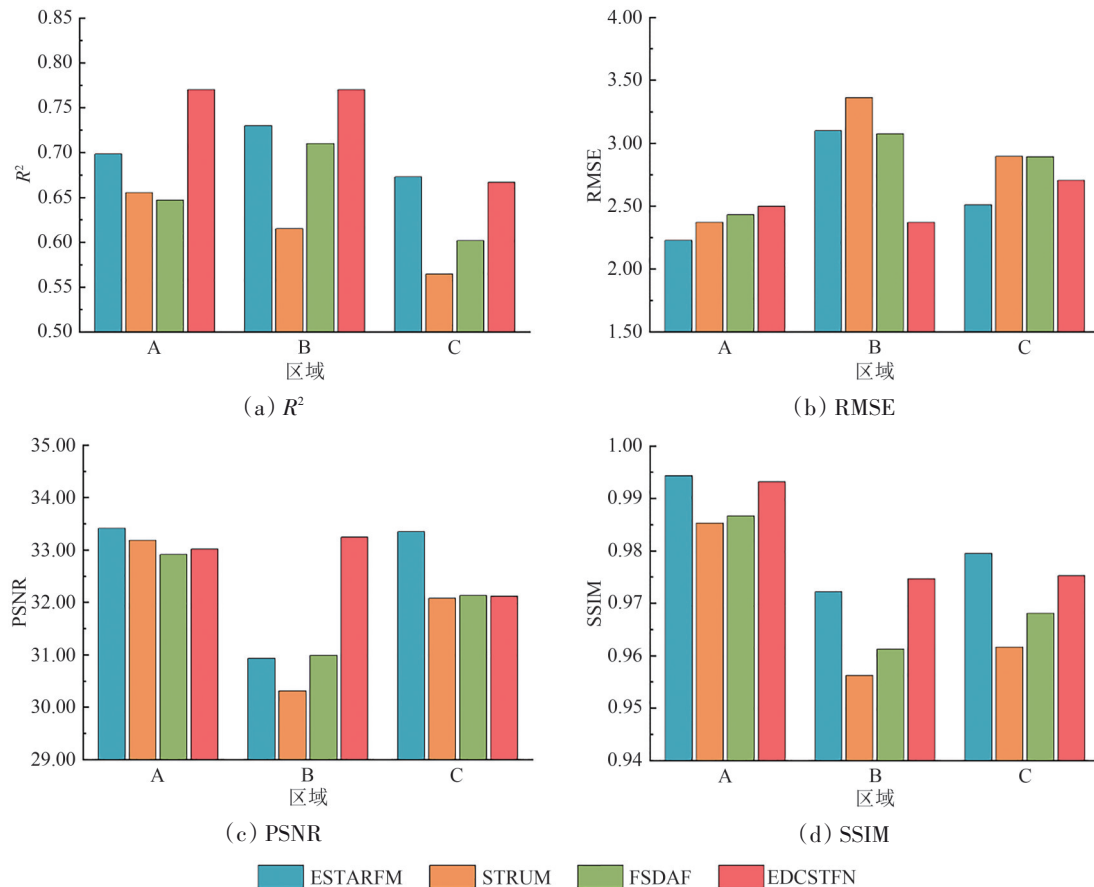


图6 不同区域4种时空融合模型融合结果的平均量化结果

Fig. 6 The average quantization results of fusion results of four spatiotemporal fusion models in different regions

根据平均量化结果显示: A区域的融合效果最理想, C区域其次, B区域效果欠佳。A区域由于地表覆盖类型相对简单, 使得地表温度变化相对规律且稳定, 4种模型均能较好地捕捉到地物边界信息与地表温度变化信息, 因此4种模型在A区域的融合精度都相对较高且差异较小。C区域尽管地表覆盖类型也相对简单, 但由于农田作物的生长周期变化导致该区域具有明显的物候变化, 且不透水地表与耕地之间的温度反差较大, 温度变化不同步, 导致C区域的融合精度相比A区域较低。此外, 使用一对参考影像的STRUM和FSDAF模型融合精度较为接近, 且相比于使用两对参考影像的ESTARFM和EDCSTFN模型精度较低, 其中ESTARFM模型在A区域和C区域中均表现最佳。B区域由于地表覆盖类型非常复杂, 不同地物类型的地表温度在时空上存在显著差异; 同时, 在地

物高度混合的区域, 低分辨率影像中多个地物类型可能会出现在同一个像元内形成混合像元, 使得模型所能捕捉到的时相变化信息减少, 从而导致融合精度相对较低。其中, EDCSTFN模型能够通过多层卷积提取复杂的空间特征和变化规律, 在混合像元的情况下, 可以更好地识别和分离不同地物类型的特征, 使其在B区域融合精度远高于其余3种模型。此外, EDCSTFN模型在地表覆盖类型复杂的B区域的RMSE甚至优于A区域和C区域, 一方面可能与B区域参与模型训练的数据量更多有关, 另一方面由于在地表类型相对简单的区域地表温度变化可能更接近线性或弱非线性关系, EDCSTFN对复杂非线性关系的建模优势并不明显, 且模型对于简单的特征也存在过拟合的风险, 导致融合效果欠佳。

### 4.3 空间尺度差异对融合结果的影响

本节采用3个研究区测试日期30—120 m、30—240 m、30—480 m、30—960 m、120—960 m、240—960 m和480—960 m 7种空间分辨率组合的LST影像对进行融合实验,选择参考影像时以图3

的分组为依据,并以实际Landsat LST影像作为真值对融合结果进行精度评价。图7和图8分别为不同空间尺度的低分辨率影像和不同空间尺度的高分辨率影像下4种模型在3个测试日期融合结果的平均量化结果。

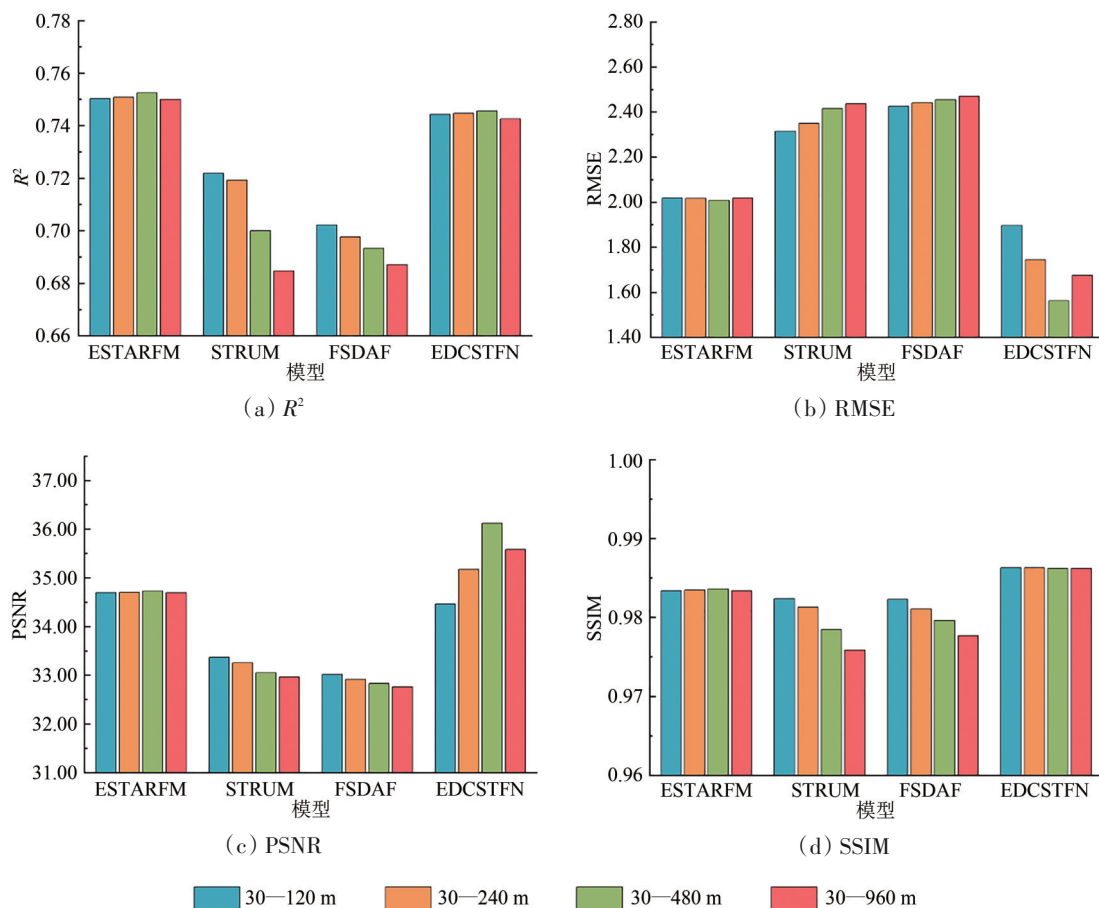
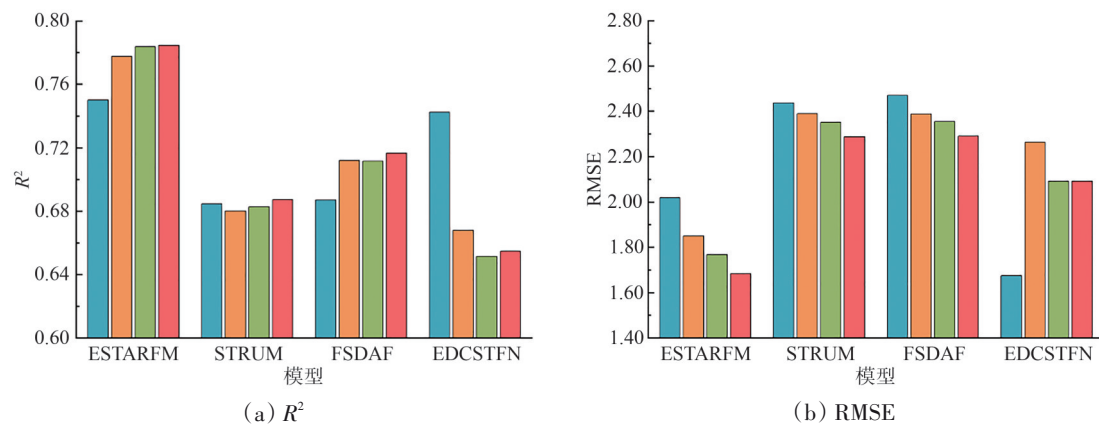


图7 不同空间尺度的低分辨率影像下4种时空融合模型在3个测试日期融合结果的平均量化结果

Fig. 7 The average quantization results of fusion results of four spatiotemporal fusion models on three test dates using low resolution images at different spatial scales



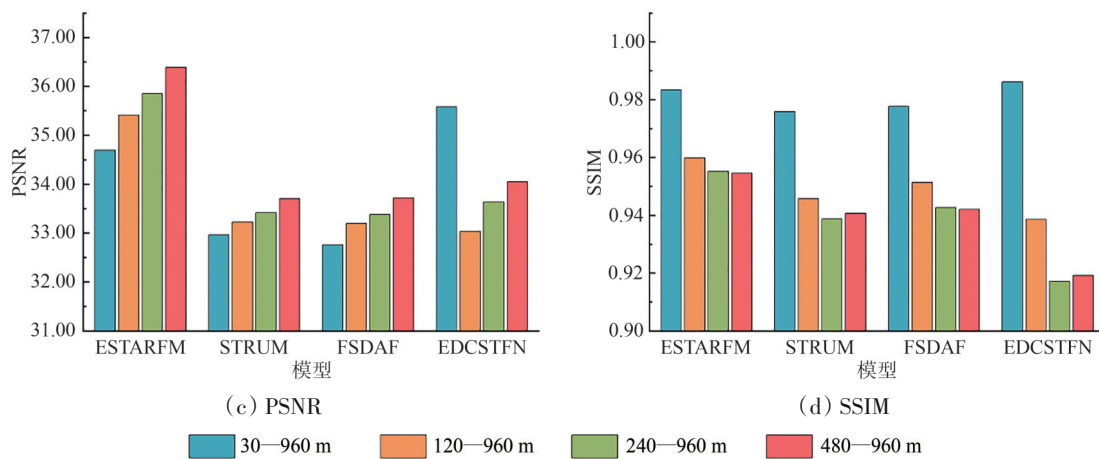


图8 不同空间尺度的高分辨率影像下4种时空融合模型在3个测试日期融合结果的平均量化结果

Fig. 8 The average quantization results of fusion results of four spatiotemporal fusion models on three test dates using high resolution images at different spatial scales

总体来看：随着空间尺度差异的变化，ESTARFM、STRUM和FSDAF模型的融合结果具有一定的规律性变化，而EDCSTFN模型虽然没有较明显的规律，但整体变化趋势与其余3种模型相反。根据图7所示：当参考影像对中高分辨率影像的空间分辨率固定，随着低分辨率影像的空间分辨率逐渐降低，STRUM和FSDAF模型由于在进行混合像元分解时对低分辨率影像的质量要求较高，导致其融合精度逐渐下降，而ESTARFM模型相对比较稳定。根据图8所示：当参考影像对中低分辨率影像的空间分辨率固定，随着高分辨率影像分辨率逐渐降低，3种模型的 $R^2$ 、RMSE和PSNR指标的精度表现为规律性上升，其中ESTARFM模型上升幅度最大，而SSIM指标的精度随着空间尺度差异的缩小反而呈现下降趋势。这是由于当高分辨率影像的空间分辨率降低时，影像中的细节信息减少，使得影像变得更加平滑和均匀，这对于全局误差（ $R^2$ 、RMSE）和噪声水平（PSNR）的改善是有益的，但也因为缺乏细节信息（如纹理和边缘），使得SSIM指标精度下降。EDCSTFN模型由于在模型训练时需要使用对应空间尺度组合的数据进行训练，因此预测不同空间尺度组合时模型使用的训练数据并不一致，导致其融合结果并没有随着空间尺度差异的变化而呈现规律性变化，但从变化趋势上可以看出EDCSTFN模型的融合精度并不会随着尺度差异的增大而降低。

#### 4.4 数据相关性对融合结果的影响

本节采用3个研究区测试日期30—960 m空间分辨率组合的LST影像对进行融合实验，选择参考

影像时以表1的分组为依据，并以实际Landsat LST影像作为真值对融合结果进行精度评价。图9、图10和图11为4种模型分别在3个测试日期的融合结果的量化结果。

从量化结果可知：随着数据相关性的降低，4种模型的融合精度也呈现规律性降低。EDCSTFN模型的 $R^2$ 指标相较于其余3种模型降幅最小，说明其在处理低相关性数据时预测影像在整体趋势上仍然较好，能够有效地保持预测结果与真实值之间的拟合程度。SSIM指标中ESTARFM和EDCSTFN模型在前两个测试日期（图9和图10）表现都相对比较稳定，在第3个测试日期（图11）中，当数据相关性降低幅度相对较大时，ESTARFM模型也表现出较大幅度的降低，而EDCSTFN模型依旧保持较小的降低幅度，表明其在低数据相关性条件下，仍然能够依靠深度卷积神经网络的强大特征提取能力较好地保持图像结构信息的一致性。STRUM和FSDAF模型在RMSE和PSNR指标上对数据相关性变化的依赖性更高，精度下降的幅度与数据相关性降低的幅度基本一致。这是由于STRUM和FSDAF模型在进行光谱解混时主要考虑的是不同地物类型的像元值变化，当数据相关性降低时，影像间的差异增大，导致在像素级别的预测上误差增大，其误差直观地反映在RMSE和PSNR指标上。而ESTARFM和EDCSTFN模型由于使用两对参考影像进行融合，更丰富的时空变化信息在一定程度上能够减轻对于数据相关性的依赖，但当信息冗余时也会引入一些不确定的偏差，导致RMSE和PSNR出现骤降的情况。

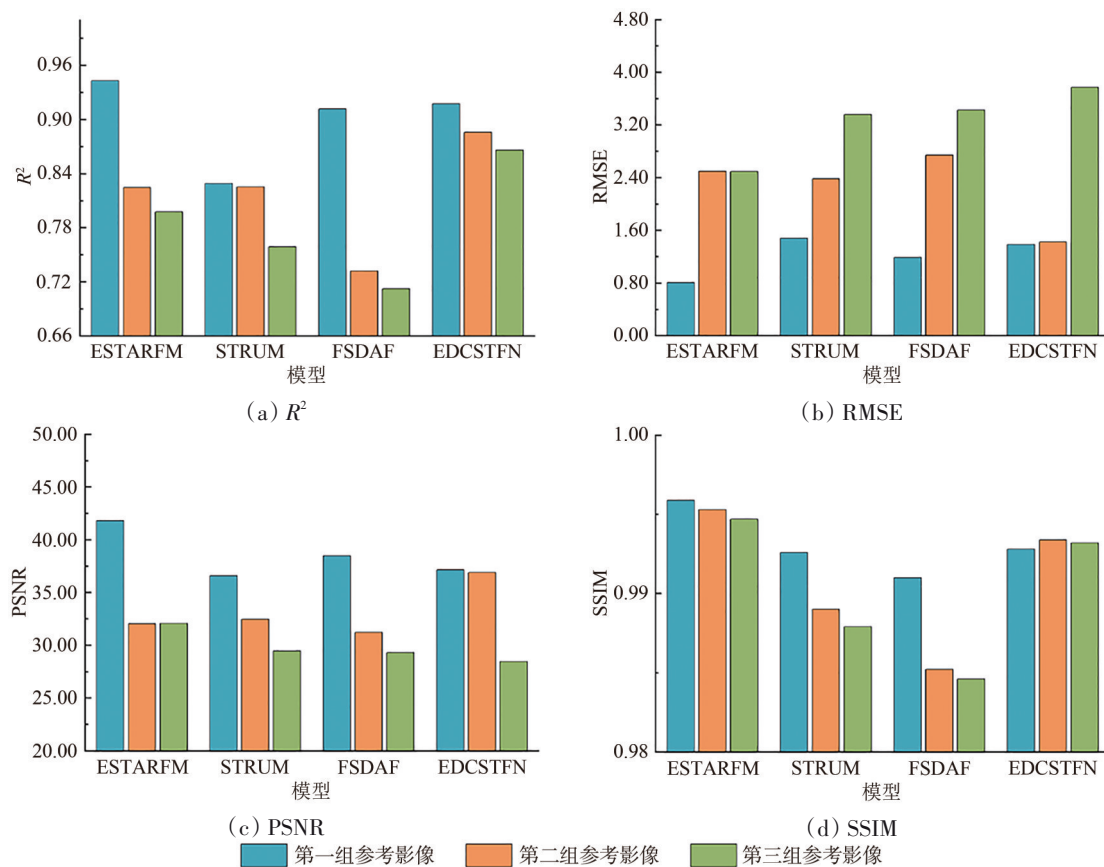


图 9 不同数据相关性下 4 种时空融合模型融合结果的量化结果(2022-12-19)

Fig. 9 Quantitative results of fusion results of four spatiotemporal fusion models with different data correlations (2022-12-19)

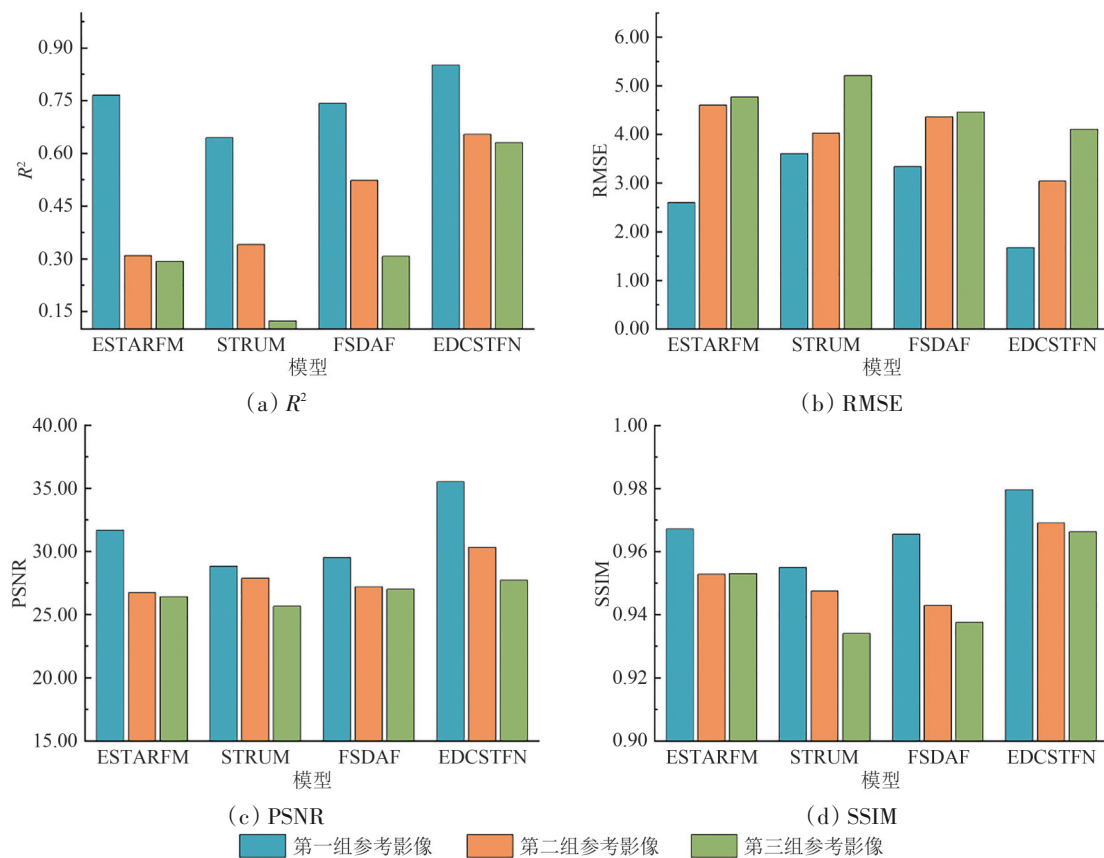


图 10 不同数据相关性下 4 种时空融合模型融合结果的量化结果(2019-09-24)

Fig. 10 Quantitative results of fusion results of four spatiotemporal fusion models with different data correlations (2019-09-24)

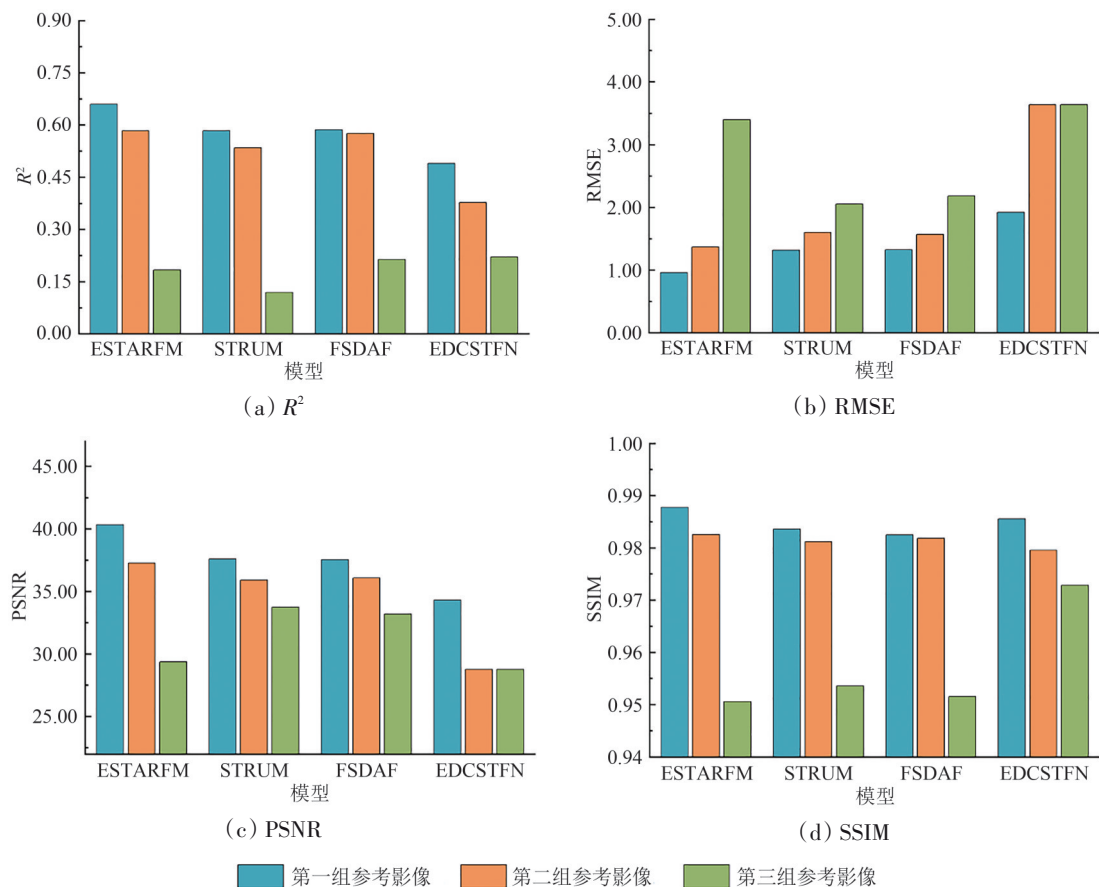


图 11 不同数据相关性下 4 种时空融合模型融合结果的量化结果(2022-11-13)

Fig. 11 Quantitative results of fusion results of four spatiotemporal fusion models with different data correlations (2022-11-13)

## 5 讨 论

目前, 遥感时空融合模型总体上分为 5 类, 包括权重模型、解混模型、学习模型、贝叶斯模型和混合模型, 本文选取的 ESTARFM、STRUM、FSDAF 和 EDCSTFN 模型分别对应权重模型、解混与贝叶斯混合模型、解混与权重混合模型以及学习模型。研究基于长时序地表温度数据集对 4 种模型的融合精度进行验证与对比分析, 并分别探究了地表环境、空间尺度差异以及数据相关性对 4 种模型融合结果的影响。

研究结果表明, EDCSTFN 模型融合 LST 精度最高, 其次是 FSDAF 和 ESTARFM 模型, STRUM 模型精度相对较差。ESTARFM 和 EDCSTFN 模型的融合影像目视效果更优, 而 STRUM 和 FSDAF 模型的融合影像整体相对平滑, 细节缺失较多。在不同地表环境条件下, EDCSTFN 模型均保持高水平精度, 尤其是在地表覆盖类型复杂的区域, 但在地物结构相对简单的区域表现不如 ESTARFM 模

型。随着空间尺度差异的增大, ESTARFM、STRUM 和 FSDAF 模型的融合精度表现为规律性下降, 而 EDCSTFN 模型并没有下降的趋势, 其中 ESTARFM 对于低分辨率数据的变化具有一定的适应能力, 但对于高分辨率数据的变化较为敏感。随着数据相关性的降低, 4 种模型的融合精度均呈现规律性降低, 其中 EDCSTFN 模型表现出相对较好的稳定性和鲁棒性。

整体来看, 学习模型在应对复杂地表环境、高空间尺度差异以及低数据相关性等场景时, 相较于传统模型具有更高的准确性与适用性。然而, 从模型易用性角度考虑, 学习模型的高精度预测很大程度上依赖于训练数据的数量和质量, 在模型训练阶段需要输入大量的高质量数据供模型学习, 但由于卫星重访周期限制以及云污染的干扰, 获取时间和空间上均保持连续的高质量地表温度数据难度较大, 尤其是在大范围区域, 极大地限制了学习模型的使用; 而传统模型对于数据的数量要求相对较少, 模型使用更加灵活, 但当

参考影像存在数据缺失时,模型也无法得到完整的融合结果。

此外,时空融合在选择参考影像时原则上选择与预测时刻影像更为接近的影像,本研究以时间跨度作为选择参考影像的衡量标准,即选择与预测时刻在时间上邻近的日期作为参考时刻。然而,地表温度数据相较于地表反射率和植被指数数据,具有更明显且不确定的波动性,线性的变化假设并不能准确反映这一现象(李冉等,2022)。结合表1的分组以及图9的实验结果也不难发现,较短的时间跨度并不一定能够提高融合精度,因此以时间跨度为标准来选择参考影像虽然是最简单且直观的,但并不是最佳的,综合多方面指标(如数据相关性)来选择参考影像可以在一定程度上提高融合精度。

通过对4种模型在不同条件下的地表温度融合结果进行分析可以发现,未来的研究可以从以下几个方面展开:

(1) 针对复杂地表环境条件下表现不佳的问题:引入地表分类信息。高精度的土地分类数据可以为模型提供更全面、综合的地表信息,从而精细地描述地表的异质性;开发基于对象的时空融合方法。相比于传统的像素级别方法,利用对象级别的特征进行时空融合可以更好地处理地表覆盖类型的复杂性,提高模型在高异质性区域的表现。

(2) 针对高空间尺度差异条件下表现不佳的问题:多源遥感数据融合。引入多种遥感数据源(如光学影像、雷达数据、高光谱数据),结合多种传感器的优势可以提供更丰富的空间和光谱信息,弥补单一数据源在高空间尺度差异下的不足;遥感影像数据增强与尺度归一化。通过多传感器融合、数据平衡等技术可以增加数据的多样性和丰富性,以此提高模型的泛化能力和鲁棒性;同时,通过尺度归一化消除不同数据源间的尺度差异,可以使地物的特征、变化和关联性等信息更易被捕捉,从而提高融合精度。

(3) 针对低数据相关性条件下表现不佳的问题:噪声过滤与降噪。在融合前对影像进行增强处理,以去除可能影响相关性的噪声和异常值,提高影像间的有效相关性;物理模型约束。将基于物理模型的约束引入到时空融合过程中,减少由于数据相关性低导致的物理不一致现象,保持

融合结果的物理合理性。

(4) 提高模型性能:引入数据立方体。云污染使研究区域无云数据的获取难度增大,因此现有研究都在尽可能减少输入影像,这导致了包含部分数据缺失的输入影像的浪费。通过构建数据立方体可以尽可能多地从时序数据中获取时空信息,从而提高模型的适用性;迁移学习。训练数据稀缺以及高计算复杂度限制了学习模型在大范围区域的应用,通过引入迁移学习技术可以简化训练过程,缓解数据不足的问题并提高模型效率。

## 6 结 论

本文选取了ESTARFM、STRUM、FSDAF和EDCSTFN共4种时空融合模型,基于Landsat 8—9和MODIS数据建立了长时序地表温度数据集,通过地面实测数据和实际Landsat LST影像数据对4种模型的融合精度进行验证与分析,并分别在不同地表环境、不同空间尺度数据组合以及不同数据相关性的条件下进行对比实验。主要结论如下:

(1) 根据地面实测数据验证结果显示,4种模型的融合LST与地面实测LST之间都具有较高的相关性,决定系数 $R^2$ 均高于0.9,其中EDCSTFN模型的融合精度最高,其次是FSDAF和ESTARFM模型,STRUM模型精度相对较差;从目视角度对比来看,ESTARFM和EDCSTFN模型的融合结果在视觉效果上与实际影像更为接近,较好地保留了地表温度分布的空间结构信息,但在细节上仍有不足,ESTARFM模型表现为部分预测值的高估,而EDCSTFN模型在温度变化较为剧烈的区域也存在部分细节丢失;STRUM和FSDAF模型的融合影像整体较为平滑,细节丢失较为严重。

(2) 4种模型在地物结构单一的区域融合效果最理想,其次是物候变化明显的区域,在地表覆盖类型复杂的区域融合效果欠佳。总体来看,ESTARFM模型在地物结构较为简单的区域预测精度最高,而在地表覆盖类型复杂的区域EDCSTFN模型更具优势,STRUM和FSDAF模型较为接近,在地物结构单一区域STRUM模型融合精度较高,而在地表覆盖类型复杂和物候变化明显的区域FSDAF模型表现更优。

(3) 随着空间尺度差异的变化,ESTARFM、STRUM和FSDAF模型的融合结果均表现出一定的规律性变化。当低分辨率影像尺度变化时,随着

空间尺度差异的增大, STRUM和FSDAF模型的融合精度逐渐降低, 而ESTARFM模型相对比较稳定; 当高分辨率影像尺度变化时, 随着空间尺度差异的减小, 3种模型的SSIM指标由于影像细节信息的减少而导致精度下降, 而 $R^2$ 、RMSE和PSNR指标的精度均表现为规律性上升, 其中ESTARFM模型的上升幅度最大, 说明其虽然对于低分辨率数据的变化具有一定的适应能力, 但对于高分辨率数据的变化较为敏感。EDCSTFN模型由于训练数据的不同导致没有较明显的规律, 但其融合精度并不会随着空间尺度差异的增大而降低。

(4) 随着数据相关性的降低, 4种模型的融合精度均呈现规律性降低。EDCSTFN模型基于深度卷积神经网络强大的特征提取和非线性建模能力使其在 $R^2$ 和SSIM指标上表现出明显的优势, 即使在数据相关性较低的情况下, 仍具有较好的鲁棒性。STRUM和FSDAF模型在RMSE和PSNR指标上对数据相关性变化的依赖性更强, 精度下降的幅度与数据相关性降低的幅度基本一致, 而ESTARFM和EDCSTFN模型由于参考影像的增加, 更丰富的时空变化信息相对减轻了对于数据相关性的依赖。

为完善本研究所使用的模型, 后续研究将从模型性能以及模型对于区域高空间异质性、数据高空间尺度差异和数据低相关性的适应能力等方面进行优化。

## 参考文献(References)

- Bateni S M and Entekhabi D. 2012. Relative efficiency of land surface energy balance components. *Water Resources Research*, 48(4): W04510 [DOI: 10.1029/2011WR011357]
- Chen B, Huang B and Xu B. 2017. Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124: 27-39 [DOI: 10.1016/j.isprsjprs.2016.12.008]
- Chen G S, Lu H L, Zou W T, Li L H, Emam M, Chen X B, Jing W P, Wang J and Li C. 2023. Spatiotemporal fusion for spectral remote sensing: a statistical analysis and review. *Journal of King Saud University-Computer and Information Sciences*, 35(3): 259-273 [DOI: 10.1016/j.jksuci.2023.02.021]
- Chen J M and Liu J. 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. *Remote Sensing of Environment*, 237: 111594 [DOI: 10.1016/j.rse.2019.111594]
- Cheng Q, Liu H Q, Shen H F, Wu P H and Zhang L P. 2017. A spatial and temporal nonlocal filter-based data fusion method. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8): 4476-4488 [DOI: 10.1109/TRGS.2017.2692802]
- Ding L R, Zhou J, Li Z L, Zhu X M, Ma J, Wang Z W, Wang W and Tang W B. 2023. Near-real-time estimation of hourly all-weather land surface temperature by fusing reanalysis data and geostationary satellite thermal infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5003918 [DOI: 10.1109/TGRS.2023.3313730]
- Duan S B, Li Z L, Gao C X, Zhao W, Wu H, Qian Y G, Leng P and Gao M F. 2020. Influence of adjacency effect on high-spatial-resolution thermal infrared imagery: implication for radiative transfer simulation and land surface temperature retrieval. *Remote Sensing of Environment*, 245: 111852 [DOI: 10.1016/j.rse.2020.111852]
- Gao F, Anderson M C, Zhang X Y, Yang Z W, Alfieri J G, Kustas W P, Mueller R, Johnson D M and Prueger J H. 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment*, 188: 9-25 [DOI: 10.1016/j.rse.2016.11.004]
- Gao F, Hilker T, Zhu X L, Anderson M, Masek J, Wang P J and Yang Y. 2015. Fusing Landsat and MODIS data for vegetation monitoring. *IEEE Geoscience and Remote Sensing Magazine*, 3(3): 47-60 [DOI: 10.1109/mg-rs.2015.2434351]
- Gao F, Masek J, Schwaller M and Hall F. 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8): 2207-2218 [DOI: 10.1109/TGRS.2006.872081]
- Gevaert C M and Garcia-Haro F J. 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sensing of Environment*, 156: 34-44 [DOI: 10.1016/j.rse.2014.09.012]
- Guo Y L, Li Y M, Zhu L, Liu G, Wang S and Du C G. 2015. An improved unmixing-based fusion method: potential application to remote monitoring of inland waters. *Remote Sensing*, 7(2): 1640-1666 [DOI: 10.3390/RS70201640]
- Huang B and Song H H. 2012. Spatiotemporal reflectance fusion via sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10): 3707-3716 [DOI: 10.1109/TGRS.2012.2186638]
- Huang B, Zhang H K, Song H H, Wang J and Song C Q. 2013. Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial-temporal-spectral earth observations. *Remote Sensing Letters*, 4(6): 561-569 [DOI: 10.1080/2150704X.2013.769283]
- Lei C Y, Meng X C and Shao F. 2021. Spatio-temporal fusion quality evaluation based on "Point" - "Line" - "Plane" aspects. *National Remote Sensing Bulletin*, 25(3): 791-802 (雷晨阳, 孟祥超, 邵枫).

2021. 遥感影像时空融合的“点”—“线” —“面”质量评价. 遥感学报, 25(3): 791-802 [DOI: 10.11834/jrs.20219334]
- Li R, Wang M M, Zhang Z J, Hu T and Liu X G. 2022. A review of spatiotemporal fusion methods for remotely sensed land surface temperature. National Remote Sensing Bulletin, 26(12): 2433-2450 (李冉, 王猛猛, 张正加, 胡添, 刘修国. 2022. 遥感地表温度产品时空融合方法研究综述. 遥感学报, 26(12): 2433-2450) [DOI: 10.11834/jrs.20210294]
- Li R B, Li H, Hu T, Bian Z J, Liu F J, Cao B, Du Y M, Sun L and Liu Q H. 2023. Land surface temperature retrieval from sentinel-3A SLSTR data: comparison among split-window, dual-window, three-channel, and dual-angle algorithms. IEEE Transactions on Geoscience and Remote Sensing, 61: 5003114 [DOI: 10.1109/TGRS.2023.3288584]
- Li S L, Wang J L, Li D C, Ran Z X and Yang B. 2021. Evaluation of Landsat 8-like land surface temperature by fusing Landsat 8 and MODIS land surface temperature product. Processes, 9(12): 2262 [DOI: 10.3390/pr9122262]
- Li X H, Wang L Y, Cheng Q, Wu P H, Gan W X and Fang L N. 2019. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. ISPRS Journal of Photogrammetry and Remote Sensing, 148: 103-113 [DOI: 10.1016/j.isprsjprs.2018.12.013]
- Li Y, Huang C L, Hou J L, Gu J, Zhu G F and Li X. 2017. Mapping daily evapotranspiration based on spatiotemporal fusion of ASTER and MODIS images over irrigated agricultural areas in the Heihe River Basin, Northwest China. Agricultural and Forest Meteorology, 244-245: 82-97 [DOI: 10.1016/j.agrformet.2017.05.023]
- Li Y T, Wu H, Li Z L, Duan S B and Ni L. 2020. Evaluation of spatiotemporal fusion models in land surface temperature using polar-orbiting and geostationary satellite data//IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. Waikoloa: IEEE: 236-239 [DOI: 10.1109/IGARSS39084.2020.9323319]
- Liu M L, Ke Y H, Yin Q, Chen X W and Im J. 2019. Comparison of five spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation. Remote Sensing, 11(22): 2612 [DOI: 10.3390/RS11222612]
- Liu S M, Li X, Xu Z W, Che T, Xiao Q, Ma M G, Liu Q H, Jin R, Guo J W, Wang L X, Wang W Z, Qi Y, Li H Y, Xu T R, Ran Y H, Hu X L, Shi S J, Zhu Z L, Tan J L, Zhang Y and Ren Z G. 2018. The Heihe Integrated Observatory Network: a basin-scale land surface processes observatory in China. Vadose Zone Journal, 17(1): 1-21 [DOI: 10.2136/vzj2018.04.0072]
- Long D, Bai L L, Yan L, Zhang C J, Yang W T, Lei H M, Quan J L, Meng X Y and Shi C X. 2019. Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. Remote Sensing of Environment, 233: 111364 [DOI: 10.1016/j.rse.2019.111364]
- Ma Y F, Liu S M, Song L S, Xu Z W, Liu Y L, Xu T R and Zhu Z L. 2018. Estimation of daily evapotranspiration and irrigation water efficiency at a Landsat-like scale for an arid irrigation area using multi-source remote sensing data. Remote Sensing of Environment, 216: 715-734 [DOI: 10.1016/j.rse.2018.07.019]
- Mao K B, Ma Y, Tan X L, Shen X Y, Liu G, Li Z L, Chen J M and Xia L. 2017. Global surface temperature change analysis based on MODIS data in recent twelve years. Advances in Space Research, 59(2): 503-512 [DOI: 10.1016/j.asr.2016.11.007]
- Meng L H, Liu H J, Zhang X L, Ren C Y, Ustin S, Qiu Z C, Xu M Y and Guo D. 2019. Assessment of the effectiveness of spatiotemporal fusion of multi-source satellite images for cotton yield estimation. Computers and Electronics in Agriculture, 162: 44-52 [DOI: 10.1016/j.compag.2019.04.001]
- Meng L H, Liu H J, Zhang X L, Xu M Y, Guo D and Pan Y. 2018. Cotton yield estimation model based on fusion image from MODIS and Landsat data//2018 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). Hangzhou: IEEE: 1-5 [DOI: 10.1109/Agro-Geoinformatics.2018.8476140]
- Peng S S, Piao S L, Zeng Z Z, Ciais P, Zhou L M, Li L Z X, Myneni R B, Yin Y and Zeng H. 2014. Afforestation in China cools local land surface temperature. Proceedings of the National Academy of Sciences of the United States of America, 111(8): 2915-2919 [DOI: 10.1073/PNAS.1315126111]
- Silva B N, Khan M and Han K. 2018. Load balancing integrated least slack time-based appliance scheduling for smart home energy management. Sensors, 18(3): 685 [DOI: 10.3390/S18030685]
- Sobrinho J A, del Frate F, Drusch M, Jiménez-Muñoz J C, Manunta P and Regan A. 2016. Review of thermal infrared applications and requirements for future high-resolution sensors. IEEE Transactions on Geoscience and Remote Sensing, 54(5): 2963-2972 [DOI: 10.1109/TGRS.2015.2509179]
- Song L S, Ding Z H, Kustas W P, Xu Y H, Zhao G L, Liu S M, Ma M G, Xue K J, Bai Y and Xu Z W. 2022. Applications of a thermal-based two-source energy balance model coupled to surface soil moisture. Remote Sensing of Environment, 271: 112923 [DOI: 10.1016/j.rse.2022.112923]
- Tan Z Y, Di L P, Zhang M D, Guo L Y and Gao M L. 2019. An enhanced deep convolutional model for spatiotemporal image fusion. Remote Sensing, 11(24): 2898 [DOI: 10.3390/rs11242898]
- Tang Y J, Wang Q M and Atkinson P M. 2023. Filling then spatio-temporal fusion for all-sky MODIS land surface temperature generation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16: 1350-1364 [DOI: 10.1109/JSTARS.2023.3235940]
- Tang Y J, Wang Q M, Tong X H and Atkinson P M. 2024. Generation of 100-m, hourly land surface temperature based on spatio-temporal fusion. IEEE Transactions on Geoscience and Remote Sensing, 62: 5001716 [DOI: 10.1109/TGRS.2024.3357735]
- Wang M M, He C, Zhang Z J, Hu T, Duan S B, Mallick K, Li H and Liu X G. 2023. Evaluation of three land surface temperature products from Landsat series using in situ measurements. IEEE Transactions on Geoscience and Remote Sensing, 61: 5000119 [DOI:

- 10.1109/tgrs.2022.3232624]
- Wang Q M and Atkinson P M. 2018. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment*, 204: 31-42 [DOI: 10.1016/j.rse.2017.10.046]
- Wang Q M, Tang Y J, Tong X H and Atkinson P M. 2024. Filling gaps in cloudy Landsat LST product by spatial-temporal fusion of multi-scale data. *Remote Sensing of Environment*, 306: 114142 [DOI: 10.1016/j.rse.2024.114142]
- Wang T, Tang R L, Li Z L, Tang B H, Wu H, Jiang Y Z and Liu M. 2018. A comparison of two spatio-temporal data fusion schemes to increase the spatial resolution of mapping actual evapotranspiration//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 7023-7026 [DOI: 10.1109/IGARSS.2018.8517768]
- Weng Q H. 2009. Thermal infrared remote sensing for urban climate and environmental studies: methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4): 335-344 [DOI: 10.1016/j.isprsjprs.2009.03.007]
- Weng Q H, Fu P and Gao F. 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sensing of Environment*, 145: 55-67 [DOI: 10.1016/j.rse.2014.02.003]
- Wu M Q, Niu Z, Wang C Y, Wu C Y and Wang L. 2012. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing*, 6(1): 063507 [DOI: 10.1117/1.jrs.6.063507]
- Wu P H, Shen H F, Zhang L P and Götsche F M. 2015. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. *Remote Sensing of Environment*, 156: 169-181 [DOI: 10.1016/j.rse.2014.09.013]
- Wu P H, Yin Z X, Zeng C, Duan S B, Götsche F M, Ma X S, Li X H, Yang H and Shen H F. 2021. Spatially continuous and high-resolution land surface temperature product generation: a review of reconstruction and spatiotemporal fusion techniques. *IEEE Geoscience and Remote Sensing Magazine*, 9(3): 112-137 [DOI: 10.1109/MGRS.2021.3050782]
- Xian G, Shi H, Zhou Q, Auch R, Gallo K, Wu Z T and Kolian M. 2022. Monitoring and characterizing multi-decadal variations of urban thermal condition using time-series thermal remote sensing and dynamic land cover data. *Remote Sensing of Environment*, 269: 112803 [DOI: 10.1016/j.rse.2021.112803]
- Xiao J, Aggarwal A K, Duc N H, Arya A, Rage U K and Avtar R. 2023. A review of remote sensing image spatiotemporal fusion: challenges, applications and recent trends. *Remote Sensing Applications: Society and Environment*, 32: 101005 [DOI: 10.1016/j.rsa.2023.101005]
- Xie D H, Gao F and Li L Y. 2018. Assessment of predictive ability of STARFM based on different Modis-Landsat image pair date//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 7247-7250 [DOI: 10.1109/IGARSS.2018.8517833]
- Yang M, Yang G J, Chen X N, Zhang Y F and You J N. 2018. Generation of land surface temperature with high spatial and temporal resolution based on FSDAF method. *Remote Sensing for Land and Resources*, 30(1): 54-62 (杨敏, 杨贵军, 陈晓宁, 张勇峰, 尤静妮. 2018. 基于 FSDAF 方法融合生成高时空分辨率地表温度. *国土资源遥感*, 30(1): 54-62) [DOI: 10.6046/gtzyyg.2018.01.08]
- Yin T G, Inglada J and Osman J. 2012. Time series image fusion: application and improvement of STARFM for land cover map and production//Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium. Munich: IEEE: 378-381 [DOI: 10.1109/IGARSS.2012.6351559]
- Zhang B H, Zhang L, Xie D, Yin X L, Liu C J and Liu G. 2015. Application of synthetic NDVI time series blended from Landsat and MODIS data for grassland biomass estimation. *Remote Sensing*, 8(1): 10 [DOI: 10.3390/RS8010010]
- Zhang M, Lin H, Long X R and Cai Y T. 2021. Analyzing the spatiotemporal pattern and driving factors of wetland vegetation changes using 2000-2019 time-series Landsat data. *Science of the Total Environment*, 780: 146615 [DOI: 10.1016/j.scitotenv.2021.146615]
- Zhou D C, Xiao J F, Bonafoni S, Berger C, Deilami K, Zhou Y Y, Frohling S, Yao R, Qiao Z and Sobrino A J. 2018. Satellite remote sensing of surface urban heat islands: progress, challenges, and perspectives. *Remote Sensing*, 11(1): 48 [DOI: 10.3390/RS11010048]
- Zhou J X, Chen J, Chen X H, Zhu X L, Qiu Y A, Song H H, Rao Y H, Zhang C S, Cao X and Cui X H. 2021. Sensitivity of six typical spatiotemporal fusion methods to different influential factors: a comparative study for a normalized difference vegetation index time series reconstruction. *Remote Sensing of Environment*, 252: 112130 [DOI: 10.1016/j.rse.2020.112130]
- Zhou Y J, Liu T X, Batelaan O, Duan L M, Wang Y X, Li X and Li M Y. 2023. Spatiotemporal fusion of multi-source remote sensing data for estimating aboveground biomass of grassland. *Ecological Indicators*, 146: 109892 [DOI: 10.1016/j.ecolind.2023.109892]
- Zhu X L, Cai F Y, Tian J Q and Williams T K A. 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. *Remote Sensing*, 10(4): 527 [DOI: 10.3390/RS10040527]
- Zhu X L, Chen J, Gao F, Chen X H and Masek J G. 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11): 2610-2623 [DOI: 10.1016/j.rse.2010.05.032]
- Zhu X L, Helmer E H, Gao F, Liu D S, Chen J and Lefsky M A. 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment*, 172: 165-177 [DOI: 10.1016/j.rse.2015.11.016]

## Performance comparison of spatiotemporal fusion models for land surface temperature: Analysis of the effect of surface environment, spatial scale differences, and data correlation

CHEN Xinyuan, SHE Chuchu, WANG Mengmeng, ZHANG Zhengjia, LI Ran,  
LI Yangyang, LIU Xiuguo

*School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430078, China*

**Abstract:** Spatiotemporal fusion is the most effective method to combine the two attributes of high spatial resolution and high temporal resolution and is essential for producing refined remote sensing data products. However, most spatiotemporal fusion models are based on surface reflectance data and the normalized difference vegetation index and rarely applied to Land Surface Temperature (LST) data. In addition, existing evaluations of spatiotemporal fusion models of LST do not fully consider the effects of different data combinations and data correlations on model performance.

This study selects four spatiotemporal fusion models: ESTARFM, STRUM, FSDAF, and EDCSTFN. A long-term LST dataset is constructed based on Landsat 8-9 and MODIS data. The fusion accuracy of the four models is verified and analyzed using ground-truth data and actual Landsat LST imagery. Comparative experiments are conducted under different surface environments, data combinations at different spatial scales, and data correlations.

Results show that (1) among the models, EDCSTFN has the highest LST fusion accuracy, followed by FSDAF and ESTARFM; STRUM has the lowest accuracy. The fusion image of ESTARFM and EDCSTFN has a good visual effect, and the fusion image of STRUM and FSDAF model is smooth. (2) The fusion effect of the four models is excellent in the single area of the ground object structure, followed by the area with obvious phenological changes. The fusion effect is poor in the area with a complex land cover type. EDCSTFN maintains high accuracy even in different surface environments, especially in areas with complex surface cover types, but it is inferior to ESTARFM in areas with a simple surface structure. (3) With the increase in the spatial scale difference, the fusion accuracies of ESTARFM, STRUM, and FSDAF models exhibit a regular decline, whereas the fusion accuracy of EDCSTFN does not decrease. (4) With the reduction in data correlation, the fusion accuracies of the four models show a regular decline, but among the models, EDCSTFN demonstrates good stability and robustness.

Overall, the complexity of the surface cover type substantially affects the accuracy of the spatiotemporal fusion models. Among the four models, ESTARFM and EDCSTFN have the highest fusion accuracy, but EDCSTFN has better stability and robustness than ESTARFM. The results of this study can provide an important reference for subsequent research on application model selection and fusion model optimization.

**Key words:** spatiotemporal fusion, LST, ESTARFM, STRUM, FSDAF, EDCSTFN

**Supported by** National Natural Science Foundation of China (No. 61801443, 42571374); National Key Research and Development Programs of China (No. 2016YFA0600302); National Natural Science Foundation of Hubei (No. 2025AFB691)