



文献 CSTR:
32001.14.11-6035.csd.2025.0067.zh



文献 DOI:
10.11922/11-6035.csd.2025.0067.zh
数据 DOI:
10.12262/lscd.17015.2574

文献分类: 地球科学

收稿日期: 2025-04-22
录用日期: 2025-10-24
发表日期: 2025-11-24

基于 Sentinel-5P 的黄河流域痕量气体数据集 (2018–2024)

曹诗宇¹, 梁继^{1*}, 冯克庭^{2,3}, 杨莲⁴

1. 湖南科技大学, 地球科学与空间信息工程学院, 湖南湘潭 411201
2. 中国科学院地球环境研究所, 黄土科学数据中心, 西安 710061
3. 中国科学院地球环境研究所, 黄土科学全国重点实验室, 西安 710061
4. 湖南省岳阳生态环境监测中心, 湖南岳阳 414000

摘要: 本研究基于 ESA Sentinel-5P 卫星的 L2 级离线数据流和 GEE L3 级标准化数据集, 系统构建了黄河流域 2018–2024 年 CH₄、HCHO、NO₂ 和 CO 四类痕量气体的时空分布数据集 (YRB-S5P-TGD)。本数据集整合了 S5P 卫星的高频全球覆盖优势与 GEE 平台的云端计算能力, 其研究采用标准化处理流程: (1) 数据筛选与质量控制: 优先采用离线数据流, 结合质量标志过滤低置信度观测值; (2) 时空聚合与重采样: 按月/年统计垂直柱浓度均值, 通过双线性重采样将原始分辨率提升至 1 km 格网; (3) 区域裁剪与标准化输出: 基于黄河流域边界生成 GeoTIFF 格式数据集, 确保兼容主流 GIS 分析需求。本数据集通过严格统一的质量控制标准, 有效避免了再处理过程中可能引入的偏差, 支持从月际变化到年际趋势的多维分析。本数据集对黄河流域痕量气体的系统性观测资料进行了有效补充, 其高时空分辨率特性为区域大气污染的精准溯源、温室气体通量估算等研究提供了重要数据支撑。

关键词: 痕量气体; 黄河流域; Sentinel-5P; Google Earth Engine 平台; YRB-S5P-TGD

数据库 (集) 基本信息简介

数据库 (集) 名称	基于 Sentinel-5P 的黄河流域痕量气体数据集 (2018–2024)
数据通信作者	梁继 (leung@lzb.ac.cn)
数据作者	曹诗宇, 梁继, 冯克庭, 杨莲
数据时间范围	2018年7月–2024年12月
地理区域	地理范围为 32°9'32"N – 41°50'20"N, 95°53'45"E – 119°3'54"E 的黄河 流域, 包括山东省、河南省、山西省、陕西省、甘肃省、四川省、青 海省、宁夏回族自治区、内蒙古自治区等省 (区)。
空间分辨率	1 km
数据量	7.3 GB
数据格式	GeoTIFF
数据服务系统网址	https://doi.org/10.12262/lscd.17015.2574 https://www.scidb.cn/doi/10.12262/lscd.17015.2574

* 论文通信作者
梁继: leung@lzb.ac.cn

基金项目	湖南省自然资源厅科技计划项目（20230127DZ）；湖南省大学生创新训练计划项目（S144351）；湖南省环保科研课题（HBKT-2022018）。
数据库（集）组成	本数据集共计333个文件，涵盖CH ₄ 、HCHO、NO ₂ 、CO 四种痕量气体的逐年与逐月平均垂直柱浓度数据（单位：XCH ₄ : ppb; XHCHO、XNO ₂ 、XCO: mol/m ² ），并包含太阳天顶角、传感器天顶角等辅助参数以支持辐射校正与观测几何验证。数据文件按年/月分层存储，命名规则为：YRB_S5P_[气体类型]_[空间尺度]_[时间尺度]_[年份]_[月份].tif （如：YRB_S5P_NO2_1km_Annual_2020.tif，YRB_S5P_NO2_1km_Monthly_202012.tif）。

引言

黄河发源于青海省巴颜喀拉山脉，流经青海、四川、甘肃、宁夏、内蒙古、陕西、山西、河南、山东等省（区），最终在山东省东营市注入渤海，典型空间范围为 32°N–42°N，95.6°E–119.2°E。上游多以高原草甸和峡谷为主，中游流经黄土高原，下游形成冲积平原。流域内拥有丰富的自然资源、文化遗产和农业基地，是中国重要的生态屏障和经济走廊。作为“生态保护和高质量发展”重大国家战略的核心区域之一，黄河流域在生态安全、经济转型与文化复兴三大战略层面的重要地位，正推动流域向“人与自然和谐共生”的可持续模式发展，使之成为全国生态文明建设与经济高质量发展的关键标杆^[1]。在“双碳”目标全球推进的迫切背景下，系统解析该流域主要痕量气体垂直柱浓度的时空演变规律具有双重价值^[2-3]，这不仅可直观反映区域空气污染状况，也深刻映射着工业布局、能源消费及产业转型等深层次结构特征^[4-6]。通过立体监测网络可精准追踪污染物迁移转化规律，为差异化减排政策提供动态决策依据^[7]。这种科学支撑体系不仅助力黄河流域“双碳”目标的高效达成^[8-9]，其凝练的“流域治理”范式也可为全球气候治理提供宝贵的本地化实践经验。

黄河流域作为我国人口最为稠密、经济发展最为活跃的区域之一，其大气环境质量不仅关乎亿万人民的健康福祉，更是维系区域生态系统可持续发展的关键要素^[10-11]。近年研究表明，该流域 103 个城市空气质量较差的区域主要集中在中游能源基地，以鲁西城市群、河南大部分城市及晋东地区的工业源排放为主^[12]。54 个城市的空气污染高值区集中在流域中下游，低值区则普遍存在于中上游^[13]。在监测技术层面，传统地基观测网络虽能提供点状精确数据^[12-13]，但对流域尺度污染物迁移转化规律的捕捉仍存在空间覆盖不足的局限。随着卫星遥感技术的突破，Sentinel-5 Precursor（S5P）凭借其对流层观测仪（Tropospheric Monitoring Instrument, TROPOMI）传感器的高光谱分辨率（UV-VIS 270–500 nm、NIR 675–775 nm、SWIR 2305–2385 nm，光谱分辨率 0.25–1.0 nm）和每日全球覆盖能力^[14]，已在大气温室气体定量反演领域展现其时空优势^[15-16]。在黄河流域的汾渭平原，任欢欢（2023）利用 TROPOMI 传感器研究了日益严重的大气臭氧污染^[17]，Wu（2023）等人利用 TROPOMI 数据研究了 NO₂ 的高值区集中于低海拔平原及人口密集河谷^[18]。然而现有研究多聚焦于单一气体大中尺度的短期观测，且地面观测需借助多轴差分吸收光谱技术（Multi-Axis Differential Optical Absorption Spectroscopy, MAX-DOAS）和激光雷达（Light Detection and Ranging, LiDAR）等揭示垂直过程，尚未形成覆盖黄河流域全境中小尺度的多组分痕量气体时空连续数据集，制约了污染物协同效应与区域传输机制的系统解析。

因此，本研究基于 Google Earth Engine（GEE）平台，对 TROPOMI L2 级和 GEE 平台 L3 级数据的主要痕量气体（如 CH₄、HCHO、NO₄、CO 等）垂直柱浓度时空产品进行系统处理，最终生成

GeoTIFF 格式的新数据集 YRB-S5P-TGD（Sentinel-5P Trace Gas Dataset for the Yellow River Basin）。

1 数据采集和处理方法

1.1 数据集来源

1.1.1 TROPOMI L2 产品

欧洲航天局 S5P 卫星搭载 TROPOMI 传感器生成的 L2 级数据集，是大气成分监测领域具有里程碑意义的高分辨率全球产品。该数据集基于卫星 L1b 观测数据，经过严格的反演算法处理，直接提供关键大气痕量气体和气溶胶的垂直柱浓度信息，包括但不限于臭氧、二氧化氮、二氧化硫、甲醛、一氧化碳、甲烷以及气溶胶和云的特性。其卓越的空间分辨率（最初约 7 km×3.5 km，后优化至 5.5 km×3.5 km）显著提升了从太空观测城市及点源污染的能力。自 2017 年 10 月开始运行以来，TROPOMI L2 数据凭借其近全球每日覆盖的特性，已成为全球和区域空气质量监测、污染物溯源追踪、火山喷发预警、紫外线辐射评估以及气候研究不可或缺的核心数据源。

TROPOMI L2 数据采用时间分箱存储，包含观测值及其对应的经纬度地理信息，但缺乏统一的网格化处理。

1.1.2 TROPOMI L3 产品

GEE 提供格网化的 TROPOMI L3 产品，该产品仅在 GEE 平台可用。它通过调用欧空局大气工具箱（Atmospheric Toolbox）中的科学地球观测数据协调工具集（Data harmonization toolset for scientific earth observation data, HARP），将 L2 数据中的地球物理参数投影并插值到规则的时空网格，该网格确保了均匀的间距。经过校准和地理配准，为每个轨道生成独立的网格文件和 L3 级标准化大气成分离线数据流^[19]。

本研究以 GEE 平台 L3 级离线数据流为基础数据源，选取甲烷（CH₄）、甲醛（HCHO）、二氧化氮（NO₂）和一氧化碳（CO）等痕量气体的垂直柱浓度产品（XCH₄、XHCHO、XNO₄、XCO），其中变量名中的“X”明确指示了这些数据产品经过了标准化处理，代表该气体在整个垂直干空气柱中所占摩尔分数的浓度值。综合整理 GEE L3 级产品信息^[19]和 L2 产品用户手册^[20]等，汇总 L3 级痕量气体产品的基本信息如表 1 所示。

表 1 GEE S5P L3 级部分痕量气体离线数据流的原始数据信息

Table 1 Original data information of the selected trace gases in the GEE S5P Level 3 offline data stream				
痕量气体	XCH ₄	XHCHO	XNO ₂	XCO
离线数据	CH4_column_volume _mixing_ratio_dry_air	Tropospheric _HCHO_column _number_density	NO2_column _number_density	CO_column _number_density
浓度单位	ppb	mol/m ²	mol/m ²	mol/m ²
值域范围	1285–2405	-0.0172–0.0074	-0.0172–0.0074	-34.43–5.71
质量标志	qa_Value ≥ 0.50	qa_Value ≥ 0.50	qa_Value ≥ 0.75	qa_Value ≥ 0.50
VAA/deg	-180–180	-180–180	-180–180	-180–180

痕量气体	XCH ₄	XHCHO	XNO ₂	XCO
VZA/deg	1–60	0.098–66.57	0.098–66.57	1–60
SAA/deg	-180–180	-180–180	-180–180	-180–180
SZA/deg	6–70	8.76–101.17	8.76–101.17	9–80
起始时间	2019-02-08	2018-12-05	2018-12-05	2018-06-28
重访间隔	2 天	2 天	2 天	2 天
像素大小	1113.2 m	1113.2 m	1113.2 m	1113.2 m
监测波段	710–770 nm, 2314–2382 nm	328–335 nm	405–465 nm	2305–2385 nm
GEE L3 描述	https://developers.google.com/earth-engine/datasets/catalog/sentinel-5p			
L2 产品手册	https://sentiwiki.copernicus.eu/web/s5p-products#S5P-Products-L2			
PUM+ATBD	$1 \text{ mol/m}^2 = 6.022140857 \times 10^{19} \text{ molec/cm}^2 = 6.022140857 \times 10^4 \text{ Pmolec/cm}^2$			

备注：VAA / VZA: View Azimuth / Zenith Angle; SAA / SZA: Solar Azimuth / Zenith Angle; PUM: Product User Information; ATBD: Algorithm Theoretical Basis Document。

1.2 数据处理方法

1.2.1 实验设计概况

本研究基于 TROPOMI L2 级和 GEE L3 级每两日痕量气体数据产品，采用多级数据融合分析框架，构建了黄河流域大气成分时空合成数据集 YRB-S5P-TGD，其流程如图 1 所示。整个流程兼顾了数据的空间精度（裁剪、重采样）、时间分辨率（逐月/逐年合成）和质量控制（角度过滤、qa_Value、精度验证），旨在为黄河流域的空气质量监测、污染物溯源等研究提供标准化、高可信度的痕量气体数据产品。

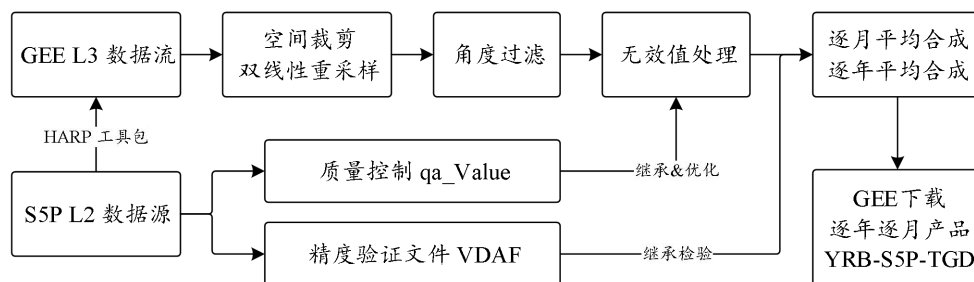


图 1 黄河流域痕量气体垂直柱浓度时空产品 YRB-S5P-TGD 的制备

Figure 1 Generation of the trace gas vertical column concentration products in the Yellow River Basin (YRB-S5P-TG)

1.2.2 HARP 工具包

HARP 主要用于输入、处理和将卫星或模型数据与相关数据进行比较。它区分了重网格化（regridding）和重新分箱（rebinning）两种关键操作。重网格化通过插值计算新网格点的值，基于邻近网格点值之间的距离进行加权平均；而重新分箱则考虑区间，根据与该区间相交的所有值的（加权）平均值为每个区间分配一个值。这两种操作与 HARP 关于轴变量的约定密切相关^[21]。

HARP 通过重网格化机制实现了 TROPOMI L2 产品从 $5.5 \text{ km} \times 3.5 \text{ km}$ 到 $1.1 \text{ km} \times 1.1 \text{ km}$ 的空间分辨率降尺度处理。该处理的核心是采用双线性插值算法，基于原始网格点与目标网格的空间位置

关系进行加权平均计算。具体实现中,通过 `bin_spatial` 函数将不规则分布的 L2 数据投影到规则的经纬度网格,并支持自定义网格分辨率,同时利用轴变量精确控制网格边界匹配,确保地理定位的准确性。

在 HARP 数据处理过程中,系统会根据变量特性采用不同的处理策略。对连续变量(如 NO_2 柱浓度)默认使用面积加权平均算法,以保持物理量的总量守恒;同时自动继承原始数据的质量标志(`qa_Value`)到新网格,保证数据质量的连续性。此外,HARP 还支持通过 GEE 平台的 Harpcovert 工具生成标准化的 L3 级数据流,有效避免了跨轨道数据聚合可能引入的误差。

值得注意的是,虽然这种降尺度处理能够改善数据的空间表现形式,但其本质上仍属于数据重采样范畴,无法突破传感器原始分辨率的物理限制。在实际应用中,特别是在处理具有强空间梯度的数据(如火山羽流)时,需要辅以地面观测数据进行验证。

1.2.3 数据预处理

依据图 1 YRB-S5P-TGD 产品的制备流程,数据的预处理包括以下步骤:

(1) 裁剪与重采样

本研究生产的痕量气体垂直柱浓度时空数据集 YRB-S5P-TGD,使用黄河流域的矢量边界^[22]对 GEE 平台 L3 级痕量气体数据集进行空间裁剪;同时,为了与其他产品如土地覆盖等保持一致的空间分辨率,将 1113.2 m 的空间分辨率经过双线性方法重采样到 1000 m。

双线性重采样是一种平衡精度与效率的空间数据处理方法。该方法通过四个邻近网格点的距离加权计算,在保持数据连续性的同时,有效减少了最近邻法的阶梯伪影和三次卷积法的过度平滑问题,能够较好地保留原始数据的空间分布特征。这一特性使其特别适合处理 NO_2 等大气成分的浓度数据。该方法可与质量标志(`qa_Value`)结合使用,实现数据缺失区域的智能插值。但双线性重采样无法突破传感器的固有分辨率限制,在处理强空间梯度特征时可能会出现轻微平滑现象。

(2) 角度过滤

大气辐射传输模型中,太阳天顶角(Solar Zenith Angle, SZA)的余弦项作为辐射传输方程的核心几何参数,决定着太阳直接辐射与天空散射辐射的定量关系。SZA 的物理效应主要体现在两个层面:①当 SZA 增大导致大气相对光学路径长度呈 $\sec(\text{SZA})$ 倍增时,太阳辐射与大气组分的相互作用时间显著延长,引发更强烈的大气分子散射和气溶胶吸收,使得地表接收的直接辐射呈指数衰减^[23]。统计表明,在晴空无云条件下,SZA 对直接辐射方差的解释度超过 50%,成为下行短波辐射的最敏感参数^[24]。②在散射辐射分量方面,SZA 通过调制大气有效光学厚度产生非线性调控作用,对辐射分量的调控具有显著的非对称性。具体而言,当 SZA 从 30° 增至 75° 时,地表直接辐射通量下降约 72%,而散射辐射在 SZA 超过 50° 后其占比超过 60%;当 $\text{SZA}=70^\circ$ 时,散射占比会跃升至 78%。这种角度依赖性的散射再分配在薄云条件下尤为显著,其非线性关系凸显了 SZA 在辐射分量动态平衡中的核心地位^[23]。

为提高痕量气体垂直柱浓度反演的数据集质量,同时满足大气辐射传输模型对 SZA 的临界要求,将所选痕量气体数据集的传感器天顶角和太阳天顶角均限定在 75° 以内,对不符合范围的观测值进行剔除,避免极端角度下大气辐射传输模型的不确定性对痕量气体反演精度的影响,有效降低了高散射路径下的反演误差。

(3) 无效值处理

该流程旨在平衡数据的完整性与可靠性,同时为后续的应用提供稳健的输入约束。非有效数据

通常定义为因观测条件或算法局限性导致的不可靠测量结果，主要源自云层遮蔽、气溶胶光学厚度超标或仪器噪声干扰等系统性误差。针对此类数据，通常采用两种处理策略：①在质量控制筛选中，通过质量标志 `qa_Value` 预设阈值（如 ≥ 0.5 ），剔除置信度不足的观测值；②在后处理算法校正中，对物理量异常值（如负的辐射通量）实施边界截断或统计回归修正，以确保数据集的一致性与可用性。

TROPOMI L2 和 GEE L3 级痕量气体垂直柱浓度产品中负值的出现，反映了反演算法在低信噪比或复杂场景下的局限性，可通过质量标志过滤和后处理算法校正进行综合处理。用户根据具体应用场景决定是否保留或剔除负值，以确保数据的科学性和适用性。负值通常被标记为“无效像素”，自动截断、剔除并替换为填充值（如 NaN）。当负值伴随低质量标志时，数据不可信，必须剔除；污染区域的负值可能指示后处理算法失效，需优先剔除；而清洁区域的负值可能属于合理噪声范围。在排放量估算等定量分析中，负值因缺乏物理意义应被排除。然而，若负值在验证数据的误差范围内，可视为噪声背景，但在统计时需明确标注不确定性。

在标准化预处理流程中，系统对 L3 级痕量气体每两日产品的空间缺失区域实施空白栅格化标识，以消除空间不连续性对分析结果的干扰。同时，空间裁剪、重采样及无效值剔除算法被集成至月均值合成模块，通过多尺度数据融合确保空间维度的统计一致性。

1.2.4 时空合成方法

（1）逐月合成

生成从 2018 年 7 月至 2024 年 12 月的完整月份序列，避免时间断层。当逐月遍历自适应均值合成算法时，时间尺度上有效数据覆盖度若低于预设阈值（ $\geq 30\%$ 有效天数），执行无效像元剔除策略并输出空白栅格，同时设置属性标记为缺失状态，保留原始时间戳（年月）的元数据，确保时间序列的连续性。逐月合成产品的输出变量包括目标气体垂直柱浓度、太阳天顶角和传感器天顶角。

（2）逐年合成

按自然年（1 月 1 日至 12 月 31 日）划分时间段；对每年逐月数据计算均值，若无有效数据则输出空白栅格；输出变量与逐月合成一致。

（3）聚合优化

逐月/逐年均值合成的方法，通过时间平均能有效降低某单一时期的随机噪声。对于无有效观测月份的缺失数据，将生成空值掩膜图像并标记属性，避免无效值干扰统计分析。

2 数据样本描述

黄河流域痕量气体数据集 YRB-S5P-TGD，包含 XCH_4 、 XHCHO 、 XNO_4 、 XCO 的逐月与逐年平均合成数据。数据以 GeoTIFF 格式存储，空间分辨率为 1 km，坐标系为 WGS84，时间跨度为 2018 年至 2024 年，覆盖整个黄河流域。以 2021 年 11 月的 XCH_4 数据为例，结合表 1 描述波段组成的典型样本特征，见图 2。

（1）`CH4_column_volume_mixing_ratio_dry_air`: 甲烷垂直柱体积混合比（干空气），单位为 ppb；

（2）`Solar_Zenith_Angle`: 太阳天顶角 SZA，单位为度（Deg，°），用于质量控制；

（3）`View_Zenith_Angle`: 传感器天顶角 VZA，单位为度（Deg，°），用于质量控制。

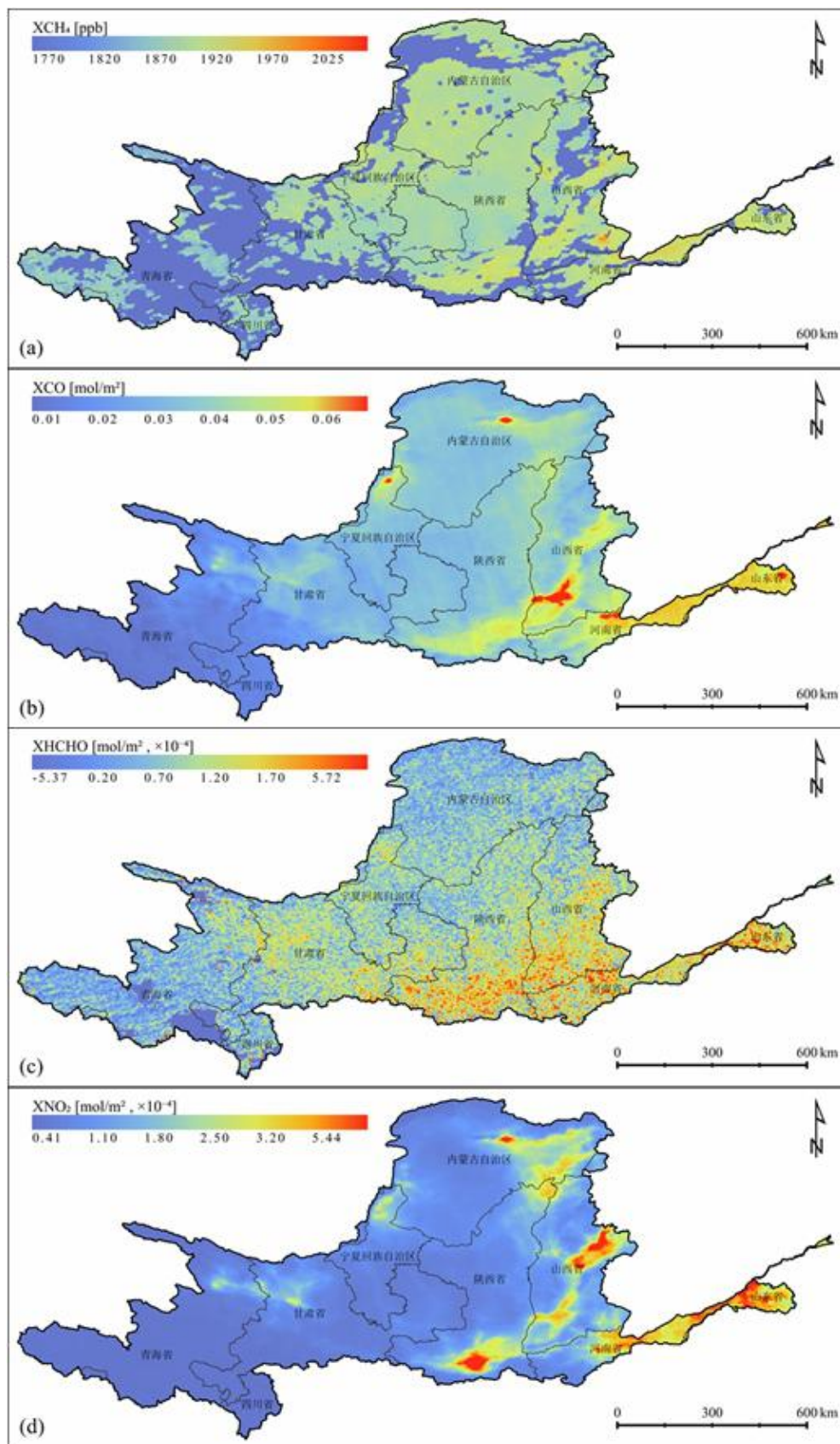


图2 2021年11月典型痕量气体垂直柱浓度(XCH_4 、 $XHCHO$ 、 XNO_2 、 XCO)的空间分布

Figure 2 Spatial distribution of typical trace gas vertical column concentrations (XCH_4 , $XHCHO$, XNO_2 , XCO) in November 2021

3 数据质量控制和评估

3.1 数据质量控制方法

在 TROPOMI L2 级数据向 L3 级产品的标准化过程中，L2 数据已经通过 HarpConvert 工具完成了云掩膜、无效像元过滤以及空间重采样等操作。GEE L3 数据流集成了官方质量控制标志 QA4ECV，通过剔除质量检查不达标的像素，提高了数据的可用性和可靠性。此外，通过量化每个像素的质量权重系数，为用户提供了一个优化数据使用的判别工具。

本研究的时空连续数据集 YRB-S5P-TGD（月尺度/年尺度），严格继承了 L3 级产品的质量筛选结果。在 L3 原始空间分辨率（1.1 km）双线性内插重采样至 1 km 的前提下，通过质量权重系数的传递机制能有效提升数据产品的科学应用价值。

3.2 数据缺失与断点说明

3.2.1 缺失数据成因

YRB-S5P-TGD 的缺失数据存有两部分原因：一是间歇性缺失，比如研究区的云覆盖率较高，导致单月有效像元不足时，对应时段输出为空值；二是传感器异常，比如 CH₄ 数据由于提供商的服务中断，2022 年 7 月 26 日至 2022 年 8 月 31 日期间没有数据。

3.2.2 断点处理

本研究采用空值检测与结构化占位的策略来处理时间断点。在数据预处理阶段，系统通过阈值响应实现月度值和年度值的精准筛除（方法细节见 1.2.3 节和 1.2.4 节）；在数据导出阶段，执行无效像元剔除策略并输出空白栅格，同时设置属性标记为缺失状态，保留原始时间戳（年月）的元数据，确保时间序列的连续性；在数据导出阶段自动跳过空值影像，既维持了时序的完整性，又避免了无效数据的干扰。

3.3 产品验证继承

S5P TROPOMI L2 级产品数据验证的核心地面网络，主要包括总碳柱观测网络（Total Carbon Column Observing Network, TCCON）^[25]、大气成分变化探测网络（Network for the Detection of Atmospheric Composition Change, NDACC）^[26]、潘多尼亚全球网络（Pandonia Global Network, PGN）^[27] 等。为全面验证 S5P 卫星 L2 级痕量气体产品（如 XCH₄、XHCHO、XNO₂、XCO 等）的精度，TCCON、NDACC 与 PGN 网络通过技术互补与数据协同，构建了覆盖总柱浓度、垂直分布、高频监测的多维度验证体系（如表 2 所示）。

表 2 三大观测网络对比与协同作用

Table 2 Comparison and synergy of three major observation networks

监测网络	TCCON ^[25]	NDACC ^[26]	PGN ^[27]
监测网络链接	https://tccodata.org/	https://ndacc.larc.nasa.gov/	https://www.pandonia-global-network.org/
核心技术	地基高分辨率傅里叶变换红外	FTIR + LiDAR + DOAS	标准化低成本的紫外-可见光谱

监测网络	TCCON ^[25]	NDACC ^[26]	PGN ^[27]
	外光谱仪（FTIR）	（MAX-DOAS + ZSL-DOAS）	仪（EnviFit）
网络特点	高精度，“黄金标准”	多技术整合，“全能平台”	高频次，广覆盖，“轻量”
垂直分辨率	中等（垂直柱浓度）	高（≤1 km，垂直结构分层）	中等（垂直柱浓度）
波长范围	总范围：0.625–2.5 μm 核心段：1.00–2.5 μm CO ₂ ：1.6 μm、2.06 μm CH ₄ ：1.65 μm 吸收带 N ₂ O：近红外弱吸收线	紫外-红外（290 nm–25 μm） FTIR：中红外 2.5–25 μm； MAX-DOAS：290–490 nm ZSL-DOAS：330–600 nm LiDAR：532 nm、1064 nm	紫外-可见光 （290–490 nm） NO ₂ ：437–450 nm 吸收带
数据精度 （相对误差）	CO ₂ ：0.1%–0.3%（400 ppm） CH ₄ ：0.2%–0.5%（1900 ppb）	O ₃ ：3%–5%；CO ₂ 、CH ₄ ：> 1% HCl：5%–8%；HNO ₃ ：7%–10%；	NO ₂ ：5%–15%；O ₃ ：8%–12% CH ₄ ：1%–2%；CO ₂ ：0.5%–1%
适用场景	碳循环与卫星校准	长期气候变化研究	城市污染实时监测
全球站点	39 个	118 个	>150 个
中国站点	2 个：香河、合肥	1 个：合肥	京津冀、长三角、珠三角
S5P 验证参数	CH ₄	NO ₂ 、CO、CH ₄ 、HCHO	NO ₂ 、O ₃
S5P 验证链接	https://mpc-vdaf.tropomi.eu/index.php		

TCCON^[25]作为温室气体高精度观测的基准网络，提供 CH₄、CO₂ 垂直柱浓度的“黄金标准”数据，直接校准 S5P 反演算法中碳气体的系统偏差；NDACC^[26]则依托其多技术融合优势，通过傅里叶变换红外光谱（Fourier Transform Infrared Spectroscopy, FTIR）技术量化平流层臭氧（O₃）对温室气体光谱的交叉干扰，利用 MAX-DOAS 解析污染区域对流层 NO₂、HCHO 的垂直廓线，并结合天顶散射光差分吸收光谱技术（Zenith-Scattered Light Differential Optical Absorption Spectroscopy, ZSL-DOAS）在清洁站点分离平流层 O₃ 信号，从而精准验证 S5P 产品的垂直分层精度。PGN^[27]作为广域补充网络，凭借低成本 EnviFit 光谱仪的分钟级高频观测能力，密集监测 NO₂、O₃ 垂直柱的日变化特征，填补传统站点在时间分辨率与空间覆盖上的空白，尤其适用于捕捉 S5P 过境间隙的污染动态。三者协同下，TCCON 提供绝对基准，NDACC 解析垂直结构与干扰校正，PGN 强化时空连续性，共同约束 S5P 数据在复杂大气条件下的不确定性。

基于欧空局 2024 年 12 月 16 日 S5P L2 数据产品验证报告^[28]，整理 L2 级离线数据流的产品验证信息如表 3 所示。鉴于研究区地理范围不包含上述三大网络的基础数据源，本研究采用跨尺度验证框架，即通过继承 L2 产品在纬度带（30°N–40°N）的辐射传输模型修正系数、先验约束参数和产品验证精度，间接表征 YRB-S5P-TGD 产品的可靠性。

表 3 所选择 S5P L2 级离线数据流的产品验证信息

Table 3 Validation information of the selected S5P TRPPOMI Level 2 data products

数据集	XCH ₄	XHCHO	XNO ₂	XCO
浓度单位	ppb	mol/m ²	mol/m ²	mol/m ²
算法升级	新增随机森林云过滤； 引入反照率参数优化； 后验去条纹技术修正轨	辐射传输模型（尤其反 照率参数）的优化； 垂直平滑的优化。	云产品更新；动态气溶 胶校正；先验剖面调整 与后验去条纹技术；修	SEOM-IAS 光谱数据库； 后验去条纹技术；引入 反照率校正；

数据集	XCH ₄	XHCHO	XNO ₂	XCO
	道条带噪声。		正平流层分离算法。	优化季节性偏差。
任务要求	$D \leq 1.5\%$, $\sigma \leq 1\%$	$D \leq 40\%–80\%$ $\sigma \leq 12 \text{ Pmolec/cm}^2$	对流层: $D \leq 50\%$ $\sigma \leq 0.7 \text{ Pmolec/cm}^2$	$D \leq 15\%$, $\sigma \leq 10\%$
验证集	TCCON NDACC	NDACC FTIR, MAX-DOAS, PGN	MAX-DOAS, ZSL-DOAS, PGN	TCCON NDACC
平均偏差 (Bias: D)	偏差校正产品: TCCON: +0.22% NDACC: -0.06%	FTIR: 低浓度: +30% 高浓度: -30% PGN: -33% MAX-DOAS: -41%	对流层: 污染区: -47% 清洁区: +26% ZSL-DOAS: -3.1% PGN: -8.5%	NDACC: +2.0% TCCON: -2.5%
标准方差 (σ)	偏差校正产品: TCCON: 0.75% NDACC: 0.90%	FTIR: 清洁区: 9 Pmolec/cm ² ; 污染区: 29 Pmolec/cm ² MAX-DOAS: 2–4 Pmolec/cm ²	对流层: 3.2 Pmolec/cm ² 平流层: 0.30 Pmolec/cm ² 垂直柱: 1.40 Pmolec/cm ²	去条纹后产品: NDACC: 8% TCCON: 5%

备注: NO₂ 和 HCHO 的清洁区或低浓度区: $< 2.5 \text{ Pmolec/cm}^2$; 污染区或高浓度区: $> 8 \text{ Pmolec/cm}^2$ 。

4 数据价值

黄河流域大气成分专题数据集 YRB-S5P-TGD 的核心价值主要体现在两方面：一是有效弥补了区域高分辨率大气成分监测数据的不足。本数据集以 1 km 的空间分辨率，显著提升了黄河流域大气成分的细节捕捉与表征能力，为区域精细化环境分析与政策制定提供了关键数据基础；二是在方法层面实现了流程标准化与输出创新。数据集采用统一规范的输出格式，支持与多源数据的无缝集成；通过系统的时间平滑处理增强了数据的统计稳健性，其生成的月度和年度均值有利于挖掘长期演变趋势；同时，额外提供的太阳天顶角与传感器天顶角数据，进一步拓展了多角度精细化分析的潜力，为大气遥感深度应用建立了标准化流程。

作为中国重要生态屏障区，黄河流域碳收支评估对“双碳”目标实施具有关键意义。本数据集通过融合多组分大气参数，可支持：1）重点行业温室气体排放清单验证；2）臭氧与前体物的光化学生成机制解析；3）跨境传输通量估算等研究。集成优势使 YRB-S5P-TGD 能支撑跨学科研究的不同需求，在大气成分动态追踪、污染源解析及气候驱动因子反演等场景中的应用，为黄河流域的环境治理提供了标准化、可复用的数据基准。

作者分工职责

曹诗宇（2004—），女，江西省上饶市人，在校学生，研究方向为大气环境遥感。主要承担工作：主导数据集全链条生产技术的实现、数据论文核心内容的构建、初稿的撰写。

梁继（1976—），女，湖南省岳阳市人，理学博士，副教授，研究方向为大气环境遥感。主要承担

工作：设计多级数据集融合方案、构建质量控制体系、优化数据论文框架。

冯克庭（1980—），男，宁夏海原县人，硕士，正高级工程师，研究方向为科学数据共享、地学大数据分析应用等。主要承担工作：制定数据共享策略、拓展数据集的应用价值。

杨莲（1981—），女，湖南省岳阳市人，研究生，高级工程师，研究方向为环境质量评价。主要承担工作：数据论文的框架讨论与环境应用深化。

参考文献

- [1] 中共中央、国务院. 黄河流域生态保护和高质量发展规划纲要[EB/OL]. (2021–10–08) [2025–04–14]. https://www.gov.cn/zhengce/2021-10/08/content_5641438.htm. [The Central Committee of the Communist Party of China and the State Council. Outline of the Plan for Ecological Protection and High-Quality Development of the Yellow River Basin [EB/OL]. (2021–10–08) [2025–04–14]. https://www.gov.cn/zhengce/2021-10/08/content_5641438.htm]
- [2] 刘良云, 陈良富, 刘毅, 等. 全球碳盘点卫星遥感监测方法、进展与挑战[J]. 遥感学报, 2022, 26(02): 243–267. DOI:10.11834/jrs.20221806. [LIU L Y, CHEN L F, LIU Y, et al. Satellite remote sensing for global stocktaking: methods, progress and perspectives [J]. National Remote Sensing Bulletin, 2022, 26(2): 243–267. DOI:10.11834/jrs.20221806.]
- [3] 刘良云, 陈良富, 尹增山, 等. 全球温室气体遥感监测和盘点研究综述与展望[J/OL]. 科学通报, 2025. DOI:10.1360/TB-2024-1033. [LIU L Y, CHEN L F, YIN Z S, et al. Review and perspective of remote sensing research on global greenhouse gas monitoring and stocktaking [J/OL]. Chinese Science Bulletin, 2025. DOI:10.1360/TB-2024-1033.]
- [4] RAHMAN M M, SULTANA N, VELAYUTHAM E. Renewable energy, energy intensity and carbon reduction: Experience of large emerging economies [J]. Renewable Energy, 2022, 184, 252–265. DOI:10.1016/j.renene.2021.11.068.
- [5] 金都, 刘新平. 中国能源结构转型：碳减排效应、驱动因素及碳达峰探索[J]. 安全与环境学报, 2024, 24(12): 4908–4916. DOI:10.13637/j.issn.1009-6094.2024.0074. [JIN D, LIU X P. Exploring the carbon emission reduction effects, drivers, and peak carbon strategies in China’ s energy structure transformation[J]. Journal of Safety and Environment, 2024, 24(12): 4908–4916. DOI:10.13637/j.issn.1009-6094.2024.0074.]
- [6] YE X X, LAUVAUX T, KORT E A, et al. Constraining fossil fuel CO₂ emissions from urban area using OCO-2 observations of total column CO₂ [J]. Journal of Geophysical Research: Atmospheres, 2020, 125(8). DOI:10.1029/2019jd030528.
- [7] 胡诚, 张俊清, 刘慧丽, 等. 城市尺度温室气体排放反演研究进展及发展趋势[J]. 中国环境科学, 2024, 44(12): 7036–7045. DOI:10.19674/j.cnki.issn1000-6923.2024.0236. [HU C, ZHANG J Q, LIU H L, et al. Research Progress and Development Trends of Greenhouse Gas Emission Inversion at the Urban Scale [J]. China Environmental Science, 2024, 44(12): 7036–7045. DOI:10.19674/j.cnki.issn1000-6923.2024.0236.]
- [8] 罗林聪, 丁跃清, 邢鑫, 等. 碳达峰碳中和目标下碳计量发展现状、挑战及对策建议[J]. 计量与测试技术, 2024, 50(12): 11–15. DOI:10.15988/j.cnki.1004-6941.2024.12.004. [LUO L C, DING Y Q,

- XING X, et al. Development status, challenges and recommendations for carbon measurement in the context of emission peak and carbon neutrality goals[J]. *Journal of Metrology & Measurement Technique*, 2024, 50(12): 11-15. DOI:10.15988/j.cnki.1004-6941.2024.12.004.]
- [9] JIANG F, WANG H W, CHEN J M, et al. Nested atmospheric inversion for the terrestrial carbon sources and sinks in China[J]. *Biogeosciences*, 2013, 10(8), 5311-5324. DOI:10.5194/BG-10-5311-2013.
- [10] 朱毅翔, 阚海东. 大气环境影响人群健康的研究进展与展望[J]. *兰州大学学报(医学版)*, 2025, 51(01): 1-11, 25. DOI:10.13885/j.issn.1000-2812.2025.01.001. [ZHU Y X, KAN H D. Current research status and advances in the impact of atmospheric environment on human health[J]. *Journal of Lanzhou University (Medical Sciences)*, 2025, 51(01): 1-11, 25. DOI: 10.13885/j.issn.1000-2812.2025.01.001.]
- [11] XIA X, MENG X, LIU C, et al. Associations of long-term nitrogen dioxide exposure with a wide spectrum of diseases: A prospective cohort study of 0.5 million Chinese adults [J]. *The Lancet Public Health*, 2024, 9(12), e1047–e1058. DOI:10.1016/s2468-2667(24)00264-0.
- [12] JIANG W, GAO W D, GAO X M, et al. Spatio-temporal heterogeneity of air pollution and its key influencing factors in the Yellow River Economic Belt of China from 2014 to 2019 [J]. *Journal of Environmental Management*, 2021, 296: 113172. DOI:10.1016/j.jenvman.2021. 113172.
- [13] MA D L, ZHANG Q, HUANG Q J, et al. Analysis of temporal and spatial changes in urban air pollution and influencing factors from 2017 to 2021: a case study of 54 cities in the Yellow River Basin [J]. *Environment, Development and Sustainability*, 2024. DOI: 10.1007/s10668-024-05309-7.
- [14] VEEFKIND J P, ABEN I, MCMULLAN K, et al. TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications [J]. *Remote Sensing of Environment*, 2012, 120: 70–83. DOI:10.1016/j.rse.2011.09.027.
- [15] LORENTE A, BORSORFF T, BUTZ A, HASEKAMP O, et al. Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements [J]. *Atmospheric Measurement Techniques*, 2021, 14(1): 665–684. DOI:10.5194/amt-14-665-2021.
- [16] SADAVARTE P, PANDEY S, MAASAKKERS J D, et al. Methane emissions from superemitting coal mines in Australia quantified using TROPOMI satellite observations [J]. *Environmental Science & Technology*, 2021, 55(24): 16573–16580. DOI:10.1021/acs.est.1c03976.
- [17] 任欢欢. 汾渭平原臭氧的时空分布特征及敏感性分析[D]. 西安交通大学, 2023. DOI:10.27394/d.cnki.gxaju.2023.000003. [REN H H. Spatiotemporal Distribution Characteristics and Sensitivity Analysis of Ozone in Fenwei Plain [D]. Xi'an Jiaotong University, 2023. DOI:10.27394/d.cnki.gxaju.2023.000003.]
- [18] WU Y R, LIU H L, LIU S Y, et al. Estimate of near-surface NO₂ concentrations in Fenwei Plain, China, based on TROPOMI data and random forest model [J]. *Environmental Monitoring and Assessment*, 2023, 195(11): 1379. DOI:10.1007/s10661-023-11993-1.
- [19] Google Earth Engine (GEE) Data Catalog, Sentinel-5P L3 datasets [EB/OL]. [2025-04-14]. <https://developers.google.com/earth-engine/datasets/catalog/sentinel-5p>

- [20] Sentinel-5P Products Overview, S5P-Products-L2 information: PUM, ATBD, IODD and PRF [EB/OL]. [2025-04-14]. <https://sentiwiki.copernicus.eu/web/s5p-products#S5P-Products-L2>.
- [21] HARP. HARP Algorithms [EB/OL]. [2025-08-12]. <https://stcorp.github.io/harp/doc/html/algorithms/regridding.html>.
- [22] 国家科技基础条件平台—国家地球系统科学数据中心. 黄河流域 1:100 万上中下游分区空间界线数据集 (2020 年) [DS/OL]. DOI:10.11971/geo.2023.049.db.lwdc. <http://www.geodata.cn>. [National Earth System Science Data Center, National Science & Technology Infrastructure of China, Spatial Boundary Dataset of Upper-Middle-Lower Reaches Zoning in the Yellow River Basin (1:1,000,000) (2020) [DS/OL]. DOI:10.11971/geo.2023.049.db.lwdc. <http://www.geodata.cn>].
- [23] RICCHAZZI P, YANG S R, GAUTIER C, et al. A research and teaching software tool for plane-parallel radiative transfer in the earth's atmosphere[J]. Bulletin of the American Meteorological Society, 1998, 79(10): 2101–2114. DOI:10.1175/1520-0477(1998)079<2101: sarats>2.0.co;2.
- [24] CHEN L, YAN G J, WANG T X, et al. Estimation of surface shortwave radiation components under all sky conditions: Modeling and sensitivity analysis[J]. Remote Sensing of Environment, 2012, 123: 457–469. DOI:10.1016/j.rse.2012.04.006.
- [25] TCCON. Total Carbon Column Observing Network [EB/OL]. [2025-04-14]. <https://tccondata.org/>.
- [26] NDACC. Network for the Detection of Atmospheric Composition Change [EB/OL]. [2025-04-14]. <https://ndacc.larc.nasa.gov/>.
- [27] PGN. Pandonia Global Network [EB/OL]. [2025-04-14]. <https://www.pandonia-global-network.org/>.
- [28] VDAF. Sentinel-5P Validation Data Analysis Facility, Sentinel-5P Quarterly Validation Report [EB/OL]. (2024–12–16) [2025–04–14]. <https://mpc-vdaf.tropomi.eu/index.php>.

论文引用格式

曹诗宇, 梁继, 冯克庭, 等. 基于 Sentinel-5P 的黄河流域痕量气体数据集 (2018–2024) [J/OL]. 中国科学数据, 2025, 10(4). (2025-11-24). DOI: 10.11922/11-6035.csd.2025.0067.zh.

数据引用格式

曹诗宇, 梁继, 冯克庭, 等. 基于 Sentinel-5P 的黄河流域痕量气体数据集 (2018–2024) [DS/OL]. 中国科学院地球环境研究所黄土科学数据中心, 2025. (2025-11-21). DOI: 10.12262/lscd.17015.2574.

A dataset of trace gases in the Yellow River Basin based on Sentinel-5P (2018–2024)

CAO Shiyu¹, LIANG Ji^{1*}, FENG Keting^{2,3}, YANG Lian⁴

1. School of Geoscience and Spatial Information Engineering, Hunan University of Science and Technology, Xiangtan 411201, P.R. China

2. Loess Science Data Center, Chinese Academy of Sciences, Xi'an 710061, P.R. China

3. State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, P.R. China

4. Yueyang Ecological Environment Monitoring Center, Yueyang 414000, P.R. China

*Email: leung@lzb.ac.cn

Abstract: This study systematically develops a spatiotemporal distribution dataset (YRB-S5P-TGD) of four trace gases (CH₄, HCHO, NO₂, and CO) in the Yellow River Basin from 2018 to 2024, based on ESA Sentinel-5P satellite L2 offline data streams and GEE L3 standardized datasets. The dataset integrates the high-frequency global coverage advantage of S5P satellites with the cloud computing capabilities of the GEE platform. The study adopts a standardized processing workflow: (1) Data screening and quality control: Offline data streams are prioritized, with low-confidence observations filtered using quality flags; (2) Spatiotemporal aggregation and resampling: Monthly/annual mean vertical column concentrations are calculated, and the original resolution is upscaled to a 1 km grid using bilinear resampling; (3) Regional cropping and standardized output: GeoTIFF-format datasets are generated based on the Yellow River Basin boundary to ensure compatibility with mainstream GIS analysis requirements. By implementing rigorous and unified quality control standards, the dataset effectively avoids biases potentially introduced during reprocessing, supporting multi-scale analyses ranging from monthly variations to interannual trends. This dataset constitutes a valuable supplement to the systematic observation data of trace gases in the Yellow River Basin. Its high spatiotemporal resolution delivers critical data support for research such as precise source apportionment of regional air pollution and greenhouse gas flux estimation.

Keywords: trace gases; Yellow River Basin; Sentinel-5P; Google Earth Engine Platform; YRB-S5P-TGD

Dataset Profile

Title	A dataset of trace gases in the Yellow River Basin based on Sentinel-5P (2018–2024)
Data corresponding author	LIANG Ji (leung@lzb.ac.cn)
Data authors	CAO Shiyu, LIANG Ji, FENG Keting, YANG Lian
Time range	July 2018 – December 2024
Geographical scope	The Yellow River Basin (32°9'32"N–41°50'20"N, 95°53'45"E–119°3'54"E), covering Shandong, Henan, Shanxi, Shaanxi, Gansu, Sichuan, Qinghai, Ningxia Hui Autonomous Region, and Inner Mongolia Autonomous Region.
Spatial resolution	1 km
Data volume	7.3 GB
Data format	GeoTIFF
Data service system	https://doi.org/10.12262/lscd.17015.2574 https://www.scidb.cn/doi/10.12262/lscd.17015.2574
Sources of funding	Research Foundation of the Department of Natural Resources of Hunan Province (Grant No. 20230127DZ); Hunan Provincial College Students Innovation Training Program Project (S144351); Hunan Provincial Environmental Protection Scientific Research Project (HBKT-2022018).

Dataset composition	The dataset comprises a total of 333 files, covering monthly and annual vertical column concentration data of four trace gases: CH₄, HCHO, NO₂ and CO (units: XCH₄ in ppb; XHCHO, XNO₂ and XCO in mol/m²), along with auxiliary parameters (solar zenith angle and sensor zenith angle) to support radiometric correction and observation geometry validation. The data files are hierarchically organized by year and month, following the naming convention: YRB_S5P_[Gas]_1km_[TimeScale]_[Year][Month].tif, (e.g., YRB_S5P_NO2_1km_Annual_2020.tif, YRB_S5P_NO2_1km_Monthly_202012.tif).
---------------------	---