

# 基于三角帽法融合的中国根区土壤水分数据集 (2018–2021) 研发

田静<sup>1\*</sup>, 马红丽<sup>2\*</sup>

- 中国科学院地理科学与资源研究所陆地水循环及地表过程重点试验室, 北京 100101;
- 鄂尔多斯市河湖保护中心, 内蒙古鄂尔多斯 017100

**摘要:** 根区土壤水分 (RZSM) 是连接地表水分循环与植被生态过程的关键变量, 对中长期干旱监测、农业气象评估等具有重要指示价值。然而, 当前时空连续的根区土壤水分数据因观测困难等问题面临诸多挑战。本研究基于 2 个陆面过程模型和 3 套再分析资料的根区土壤水分数据, 采用三角帽法 (TCH) 进行融合, 生成 2018–2021 年中国区域 0–100 cm 根区土壤水分数据集。该数据集时间分辨率为日, 空间分辨率 0.25 度, 数据格式为.tif, 覆盖中国大陆区域。利用中国 2,061 个土壤水分观测站验证显示, 融合数据的 RMSE 中位数  $0.077 \text{ m}^3/\text{m}^3$ , 相关系数  $r$  中位数 0.5, Bias 峰值为 0, ubRMSE 中位数  $0.04 \text{ m}^3/\text{m}^3$ , 表明数据集具有较好可靠性与适用性, 可为区域干旱监测、生态水文分析等提供数据支撑。

**关键词:** 根区土壤水分; 三角帽法; 数据融合

**DOI:** <https://doi.org/10.3974/geodp.2025.04.04>

**CSTR:** <https://cstr.escience.org.cn/CSTR:20146.14.2025.04.04>

## 数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2025.08.08.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2025.08.08.V1>.

## 1 前言

根区土壤水分 (Root Zone Soil Moisture, RZSM) 是指植被根系主要分布区域内土壤所含的水分含量, 是连接地表水分循环与植被生态过程的关键变量。它不仅直接影响植物的水分可利用性、蒸散发、光合速率与作物产量, 还在气候系统中起着调节地气相互作用的重要作用。与表层土壤水分相比, 根区土壤水分具有更强的缓冲性和记忆性, 对中长期干旱监测、农业气象评估和生态系统建模具有更高的指示价值<sup>[1,2]</sup>。

随着气候变化驱动下干旱频发、极端水文事件加剧, 获取根区土壤水分信息对于提升

收稿日期: 2025-08-29; 修订日期: 2025-11-27; 出版日期: 2025-12-24

基金项目: 内蒙古自治区科学技术厅、鄂尔多斯市科学技术局 (ZD20232303); 国家自然科学基金 (42071327)

\*通讯作者: 马红丽, 鄂尔多斯市河湖保护中心, mhl20232023@163.com; 田静, 中国科学院地理科学与资源研究所, tianj.04b@igsrr.ac.cn

数据引用方式: [1] 田静, 马红丽. 基于三角帽法融合的中国根区土壤水分数据集 (2018–2021) 研发[J]. 全球变化数据学报, 2025, 9(4): 397–404. <https://doi.org/10.3974/geodp.2025.04.04>. <https://cstr.escience.org.cn/CSTR:20146.14.2025.04.04>.

[2] 田静. 中国根区 (0–100 cm) 土壤水分 0.25°日频数据集 (2018–2021) [J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.08.08.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2025.08.08.V1>.

干旱监测精度、指导农业灌溉决策、评估生态系统韧性具有重要意义。然而，由于观测困难、遥感穿透深度有限、模型不确定性大，当前根区土壤水分数据仍面临诸多挑战<sup>[3,4]</sup>。因此，研发融合多源数据（如遥感、气象驱动、地面观测与机器学习等）的根区土壤水分数据集，是推动水文-生态-农业综合研究的重要基础工作。

本研究基于5个陆面过程模型生成的根区土壤水分数据和再分析资料的根区土壤水分数据，利用三角帽法（Three-Corned Hat, TCH）进行数据融合，生成了中国大陆区域根区土壤水分数据集(0–100 cm)，为更好地服务于区域尺度干旱监测、生态水文过程分析及农业水分管理等应用奠定了数据基础。

## 2 数据集元数据简介

《中国根区（0–100 cm）土壤水分 0.25°日频数据集（2018–2021）》<sup>[5]</sup>的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表1。

表1 《中国根区（0–100 cm）土壤水分 0.25°日频数据集（2018–2021）》元数据简表

条目	描述
数据集名称	中国根区（0–100 cm）土壤水分 0.25°日频数据集（2018–2021）
数据集短名	RZSM_China_2018-2021
作者信息	田静, 中国科学院地理科学与资源研究所, tianj.04b@igsnr.ac.cn
地理区域	中国大陆
数据年代	2018–2021
时间分辨率	日
空间分辨率	0.25°
数据格式	.tif
数据量	99.4 MB（压缩后）
数据集组成	0–100cm 根区土壤水分平均值
基金项目	内蒙古自治区科学技术厅、鄂尔多斯市科学技术局（ZD20232303）；国家自然科学基金（42071327）
数据计算环境	Python
出版与共享服务平台	全球变化科学研究数据出版系统 <a href="http://www.geodoi.ac.cn">http://www.geodoi.ac.cn</a>
地址	北京市朝阳区大屯路甲11号100101, 中国科学院地理科学与资源研究所
数据共享政策	(1)“数据”以最便利的方式通过互联网系统免费向全社会开放, 用户免费浏览、免费下载; (2) 最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源; (3) 增值服务用户或以任何形式散发和传播(包括通过计算机服务器)“数据”的用户需要与《全球变化数据学报》(中英文)编辑部签署书面协议, 获得许可; (4) 摘取“数据”中的部分记录创作新数据的作者需要遵循10%引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的10%, 同时需要对摘取的数据记录标注数据来源 <sup>[6]</sup>
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

## 3 数据研发方法

本研究基于2个陆面过程模型生成的根区土壤水分数据和3套再分析资料的根区土壤

水分数据（表2），采用三角帽法（TCH）进行数据融合，构建了新的中国区域0–100 cm根区土壤水分数据产品，并结合中国区域2,000多个土壤水分观测站的数据进行了系统验证。研究方法主要包括两部分：

（1）0–100 cm根区土壤水分的计算：通过对各模型输出和站点观测的多层土壤水分数据进行加权平均获得。加权系数按照相邻两个土层中心深度之间的距离在总深度（100 cm）中的占比计算，体现各土层对根区水分的贡献。该加权平均过程分别独立应用于每个模型产品和站点观测数据，最终得到可比的0–100 cm根区土壤水分估算值。

（2）融合产品的构建：在获得5个模型的0–100 cm根区土壤水分产品后，采用TCH方法对它们进行误差方差估计与权重分配，生成一个无须依赖参考“真值”的融合产品。该方法能够有效评估各数据源的相对误差水平，并在此基础上优化加权方案，从而提高融合产品的综合精度和稳定性。融合结果不仅整合了多源模型的优势，也在一定程度上降低了单一数据源的不确定性。

另外，所有数据都采用算术平均法计算日均值，并采用双线性内插法重采样至0.25度，以保持相同的时空分辨率。

表2 5个模型输出的多层土壤水分数据产品汇总表

模型名称	空间分辨率	时间分辨率 (h)	土壤水分深度 (cm)
ERA5 <sup>[7]</sup>	0.1°	3	0–7、7–28、28–100、100–289
MERRA-2 <sup>[8]</sup>	0.5°×0.625°	3	0–5、0–100
CFSR <sup>[9]</sup>	0.205°×0.204°	1	0–10、10–40、40–100、100–200
GLDAS-NOAH2.1 <sup>[10]</sup>	0.25°	3	0–10、10–40、40–100、100–200
SMAP Level 4 <sup>[11]</sup>	9 km	3	0–5、0–100

### 3.1 算法原理

#### （1）0–100 cm根区土壤水分的计算

如前所述，0–100 cm RZSM的计算采用多层土壤水分数据进行加权平均获得，加权系数按照相邻两个土层中心深度之间的距离在总深度（100 cm）中的占比计算，体现各土层对根区水分的贡献。以表2中GLDAS-NOAH2.1为例，其多层土壤水分在0–100 cm之间包括3个深度，0–10 cm，10–40 cm和40–100 cm，各层占总深度（100 cm）的权重即为0.1，0.3和0.6，0–100 cm RZSM的计算公式为：

$$\theta_{RZSM} = 0.1 \times \theta_{0-10\text{cm}} + 0.3 \times \theta_{10-40\text{cm}} + 0.6 \times \theta_{40-100\text{cm}} \quad (1)$$

式中， $\theta_{RZSM}$ 是0–100 cm的土壤水分， $\theta$ 是各层的土壤水分。表2中的其它模型以及验证用的观测值也是按此方法计算。

#### （2）TCH融合方法

TCH方法是一种用于评估多个数据源相对误差并进行加权融合的方法，具有无需参考真实值、适用于多源数据误差评估的显著特点。该方法通过分析不同数据源之间的两两差异，推导各自的相对误差方差，并以此为依据确定融合权重，从而在提升融合结果精度的同时，降低了对实测数据的依赖，特别适用于缺乏高密度观测或真值难以获取的环境数据融合场景。根区土壤水分实测数据通常难以获取，因此TCH方法特别适合应用。

TCH 方法由 Tavella 和 Premoli 提出<sup>[12]</sup>。其假设  $X_i$  ( $i=1,2,\dots,N$ ) 为第  $i$  种根区土壤水分产品的时间序列, 其中  $N$  为产品数量, 本文  $N=5$ 。  $X_i$  由真实值 ( $X_i$ ) 及其误差 ( $\varepsilon_i$ ) 组成, 如下式所示:

$$X_i = X_i + \varepsilon_i \quad (2)$$

式中,  $i = 1, 2 \dots, N$ 。为求解  $\varepsilon_i$ , 首先计算  $N-1$  种产品与随机选取的参考产品 ( $X_R$ ) 之间的差值, 定义为:

$$Y_i = X_i - X_R = \varepsilon_i - \varepsilon_R \quad (3)$$

式中,  $i = 1, 2 \dots, N-1$ 。  $\varepsilon_i$  与  $\varepsilon_j$  之间的协方差为:

$$r_{ij} = \frac{1}{M-1} (\varepsilon_i - \bar{\varepsilon}_i)^T (\varepsilon_j - \bar{\varepsilon}_j) \quad (4)$$

式中,  $i, j = 1, 2 \dots, N-1$ 。  $M$  为时间样本数量。  $\bar{\varepsilon}_i$  和  $\bar{\varepsilon}_j$  分别为第  $i$  种和第  $j$  种根区土壤水分产品误差的均值。上标  $T$  代表转置。相应地,  $Y_i$  与  $Y_j$  之间的协方差可表示为:

$$S_{ij} = \frac{1}{M-1} (Y_i - \bar{Y}_i)^T (Y_j - \bar{Y}_j) = r_{ij} - r_{iR} - r_{jR} + r_{RR} \quad (5)$$

式中,  $i, j = 1, 2 \dots, N-1$ 。  $r_{ij}$ 、  $r_{iR}$ 、  $r_{jR}$  和  $r_{RR}$  分别代表  $\varepsilon_i$  与  $\varepsilon_j$  之间的协方差、  $\varepsilon_i$  与  $\varepsilon_R$  之间的协方差、  $\varepsilon_j$  与  $\varepsilon_R$  之间的协方差, 以及  $\varepsilon_R$  与  $\varepsilon_R$  之间的协方差, 并由公式 4 计算得到。但式 4 无法直接求解, 因为未知数数量多于方程数量。 Galindo 和 Palacio 基于 Kuhn-Tucker 定理提出约束极小化问题, 并对其求解<sup>[13]</sup>。目标函数  $F$  和约束条件  $H$  分别为:

$$F = \frac{1}{K^2} \sum_{i < j} r_{ij}^2 \quad (6)$$

$$H = -\frac{|R|}{|S|K} < 0 \quad (7)$$

其中,

$$K = N \sqrt{\det(S)} \quad (8)$$

$$R = \{r_{ij}\}_{N \times N} \quad (9)$$

$$S = \{S_{ij}\}_{(N-1) \times (N-1)} \quad (10)$$

式中,  $R$  是误差协方差矩阵;  $S$  是和的协方差矩阵;  $K$  是单位矩阵。  $r_{ij}$  ( $i=1,2, \dots, N$ ) 即第  $i$  和第  $j$  个产品的误差协方差, 所有产品两两交互可以得到误差协方差矩阵。融合各根区土壤水分产品时所用的权重由误差协方差矩阵的逆矩阵确定, 该矩阵在估算 TCH 方法不确定性的同时, 也考虑了误差相关性。根据 Gauss-Markov 定理, 这种方法能得到具有最小方差的加权平均值:

$$W = C^{-1} \quad (11)$$

$$X_{weighted} = (J^T W J)^{-1} (J^T W X) \quad (12)$$

式中,  $C$  为通过获得的误差协方差矩阵,  $W$  为权重矩阵。  $X_{weighted}$  为权重矩阵的变换形式,  $J$  为设计矩阵, 是一个全为 1 的向量, 即  $[1, \dots, 1]^T$ 。由此获得每个产品在数据融合时的权重值。

### 3.2 技术路线

本数据集的制作主要包括如下6个步骤：（1）土壤水分数据产品预处理：对土壤水分数据产品计算平均值；对数据进行空间重采样为0.25度，使其具有统一的空间分辨率；计算0–100 cm深度的土壤水分含量。（2）站点观测土壤水分预处理：对站点观测数据计算日平均值；对站点观测数据计算0–100 cm深度的土壤水分含量。（3）基于TCH方法对预处理后的土壤水分数据产品进行误差方差计算。（4）根据计算得到的误差方差进行权重计算，以确定不同数据产品在后续融合过程中的权重。（5）利用计算得到的权重，对土壤水分数据产品进行数据融合，得到融合后的根区土壤水分数据。（6）使用站点观测的土壤水分数据进行地面观测对比验证，评估融合后数据产品的准确性（图1）。

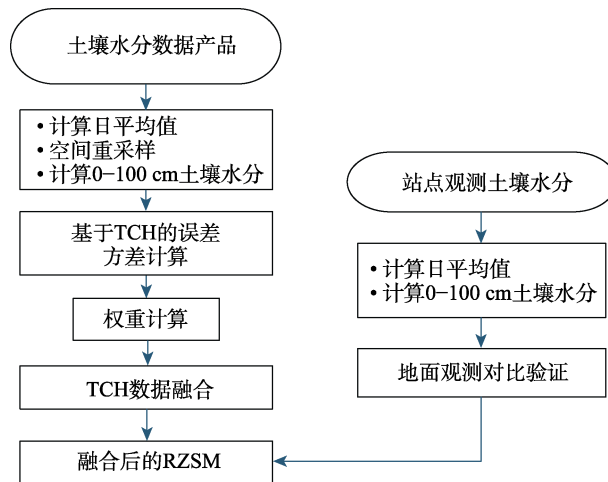


图1 数据研发技术路线图

## 4 数据结果与验证

### 4.1 数据集组成

本数据集是2018年至2021年的中国大陆区域日尺度的0–100 cm土壤水分数据，时间分辨率为日，空间分辨率为0.25°，存储为.tif格式，由1,461个数据文件组成。

### 4.2 数据结果

图2显示了中国大陆区域1–12月根区土壤水分的多年月平均值。从中可见，中国根区土壤水分呈现出显著的地域差异。总体而言，呈现出由东南向西北递减的趋势。东南地区的土壤水分含量相对较高，这主要得益于该地区丰富的降水以及相对湿润的气候条件，同时，该区域植被覆盖较好，也有助于保持土壤水分。而西北地区土壤水分含量明显偏低。西北地区深居内陆，降水稀少，气候干旱，蒸发量大，导致土壤水分难以得到有效补充。此外，青藏高原等部分高海拔地区，由于气候寒冷、冻土发育等因素，土壤水分状况也较为特殊。从时间变化特征分析，根区土壤水分存在明显的季节性波动。在春季（3–5月），随着气温逐渐回升，降水开始增加，土壤水分含量逐步上升。到了夏季（6–8月），降水更为集中，成为土壤水分的主要补给期，此时土壤水分达到全年较高水平。进入秋季（9–11月），气温逐渐下降，降水减少，土壤水分开始缓慢减少。冬季（12–2月），气温低，降水

形式多为降雪,且蒸发量小,土壤水分变化相对平稳,但整体处于较低水平。

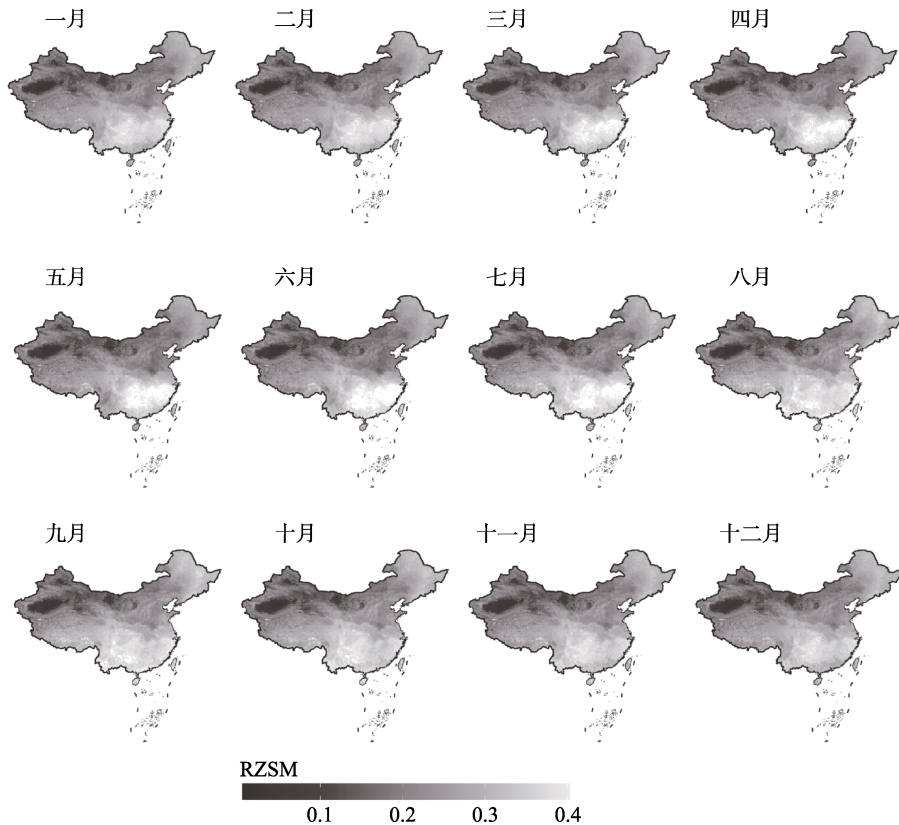


图2 中国大陆区域根区土壤水分的多年月平均值空间分布图  
(依据审图号为GS(2019)1815号的标准地图制作)

### 4.3 数据结果验证

利用全国2,061个土壤水分观测站的数据对TCH融合后的结果进行了站点验证,结果如图3和图4所示。土壤水分观测站集中分布在中东部地区,西部较稀疏。图3并没有显示出明显的空间分布规律,总体上,相关系数 $r$ 的高值主要分布在华北和南方地区。结合图4的直方图可以看出,RMSE峰值集中在 $0.05\text{--}0.10\text{ m}^3/\text{m}^3$ ,中位数为 $0.077\text{ m}^3/\text{m}^3$ ,说明多数站点误差处于中等水平,但高值长尾显示仍有部分站点误差较大。相关系数峰值在 $0.5\text{--}0.8$ ,中位数为 $0.5$ ,说明多数站点线性关联较好,但左侧低值区间仍存在少量占比。西北地区的站点普遍 $r$ 值较低。Bias峰值接近0,且大多数分布在 $-0.05\text{ m}^3/\text{m}^3$ 到 $0.05\text{ m}^3/\text{m}^3$ 之间,显示出融合结果在全国尺度上不存在明显的系统性偏差,体现了TCH方法的系统稳定性,但图3显示华北地区有低估的现象。ubRMSE峰值在 $0.03\text{--}0.05\text{ m}^3/\text{m}^3$ ,中位数为 $0.04\text{ m}^3/\text{m}^3$ ,多数站点随机误差不大,但也有少数高值区间存在。这些验证结果可见,TCH融合方法在全国尺度上具有较好的适用性和稳健性,但也仍存在改进空间,而这主要与输入的数据产品精度密切相关。值得指出的是在站点尺度对TCH融合结果进行验证时,由于观测站点的点位观测代表的是极小范围内的土壤水分状态,而融合后的数据产品为较大空间分辨率的栅格( $0.25^\circ$ ),两者在空间尺度上的不匹配可能导致较大的不确定性。

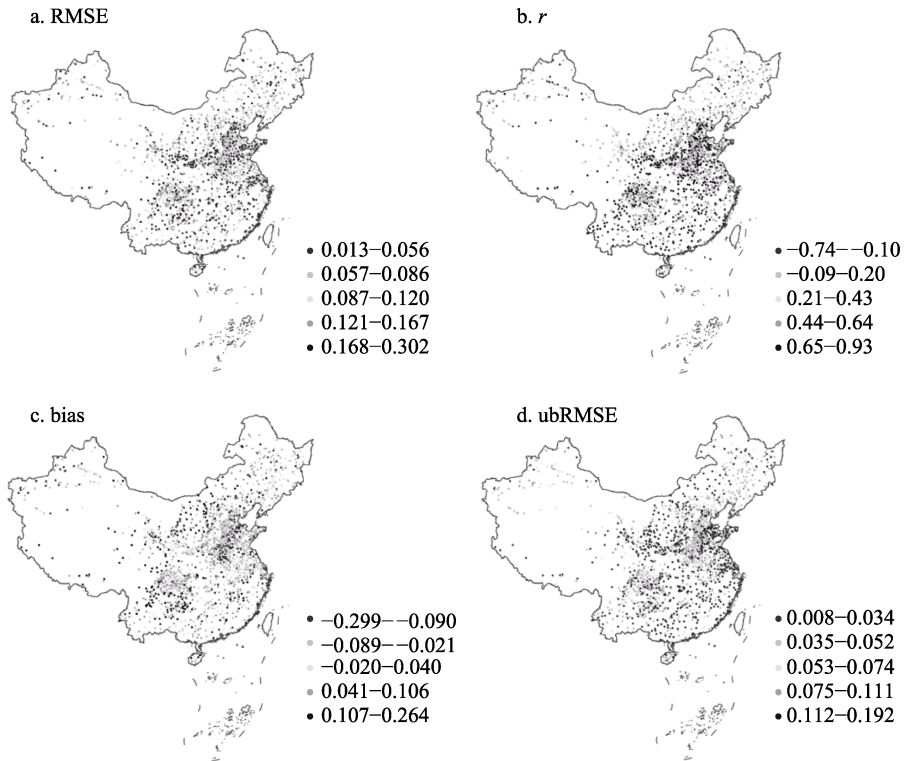


图 3 TCH 融合后中国大陆区域土壤水分数据的站点验证效果图  
（依据审图号为 GS(2019)1815 号的标准地图制作）

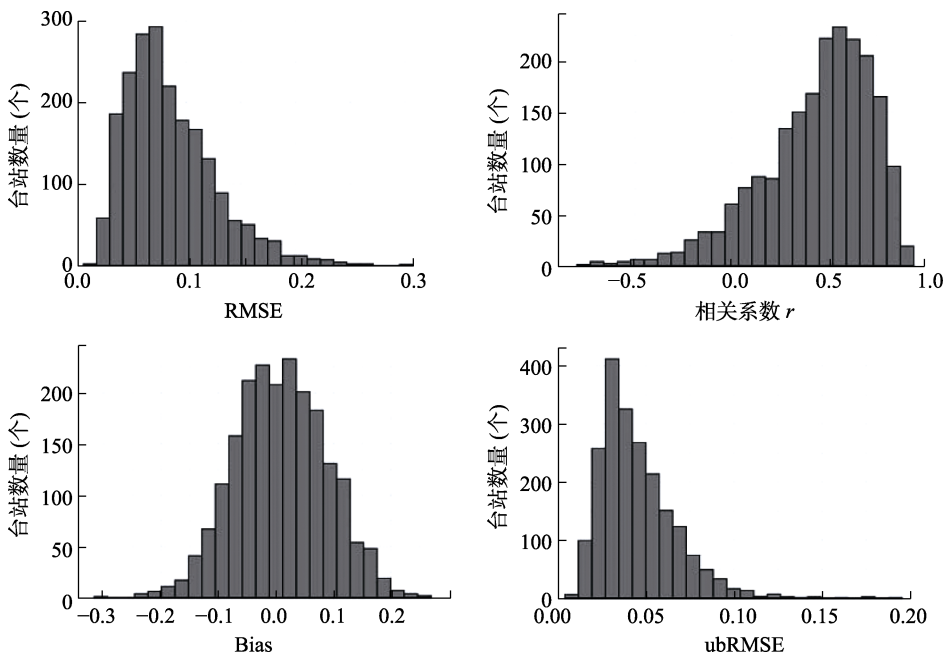


图 4 站点验证结果的直方图统计

## 5 讨论和总结

本文利用 TCH 方法对多种根区土壤水分数据进行了融合,与 2,061 个站点实测数据相比,融合后的数据产品 RMSE,  $r$ , bias 和 ubRMSE 的中位数分别是 0.077, 0.5, 0.008 和  $0.04 \text{ m}^3/\text{m}^3$ 。峰值集中在  $0.05\text{--}0.10 \text{ m}^3/\text{m}^3$ ,  $0.5\text{--}0.8 \text{ m}^3/\text{m}^3$ , 0 和  $0.03\text{--}0.05 \text{ m}^3/\text{m}^3$ 。由此说明融合后的数据具有较好的可靠性,体现了 TCH 融合方法在全国尺度上具有较好的适用性和稳健性,对于根区土壤水分这种观测难度较大,观测数据非常稀缺的地表要素尤其具有重要的应用价值。未来仍可以结合不同数据源灵活调整输入变量,提高融合精度。

**作者分工:** 马红丽对数据集的开发做了总体设计; 田静采集、处理了数据,并撰写了数据论文。

**利益冲突声明:** 本研究不存在研究者以及与公开研究成果有关的利益冲突。

## 参考文献

- [1] Tobin, K. J., Torres, R., Crow, W. T. Multi-decadal analysis of root-zone soil moisture applying the exponential filter across CONUS [J]. *Hydrology and Earth System Sciences*, 2017, 21 (9): 4403–4417.
- [2] Zohaib, M., Kim, H., Choi, M. Evaluating the patterns of spatiotemporal trends of root zone soil moisture in major climate regions in East Asia [J]. *Journal of Geophysical Research-Atmospheres*, 2017, 122 (15): 7705–7722.
- [3] Xu, L., Chen, N. C., Zhang, X., et al. In-situ and triple-collocation based evaluations of eight global root zone soil moisture products [J]. *Remote Sensing of Environment*, 2023, 254: 112248.
- [4] Tian, J., Zhang, Y. Q. Comprehensive validation of seven root zone soil moisture products at 1153 ground sites across China [J]. *International Journal of Digital Earth*, 2023, 16(2): 4008–4022.
- [5] 田静. 中国根区 (0–100 cm) 土壤水分  $0.25^\circ$  日频数据集 (2018–2021) [J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.08.08.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2025.08.08.V1>.
- [6] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [7] Bell, B., Hersbach, H., Simmons, A., et al. The ERA5 global reanalysis: preliminary extension to 1950 [J]. *Quarterly Journal of the Royal Meteorological Society*, 2021, 147(741): 4186–4227.
- [8] Reichle, R. H., Draper, C. S., Liu, Q., et al. Assessment of MERRA-2 land surface hydrology estimates [J]. *Journal of Climate*, 2017, 30(8): 2937–2960.
- [9] Saha, S., Moorthi, S., Wu, X. R., et al. The NCEP climate forecast system version 2 [J]. *Journal of Climate*, 2014, 27(6): 2185–2208.
- [10] Rodell, M., Houser, P. R., Jambor, U., et al. The global land data assimilation system [J]. *Bulletin of the American Meteorological Society*, 2004, 85(3): 381–394.
- [11] Reichle, R. H., Liu, Q., Koster, R. D., et al. Version 4 of the SMAP Level-4 soil moisture algorithm and data product [J]. *Journal of Advances in Modeling Earth Systems*, 2019, 11(10): 3106–3130.
- [12] Tavella, P., Premoli, A. Estimating the instabilities of N-Clocks by measuring differences of their readings [J]. *Metrologia*, 1994, 30(5): 479–486.
- [13] Galindo, F. J., Palacio, J. Estimating the instabilities of N correlated clocks [C]. In Proceedings of the 31th annual precise time and time interval systems and applications meeting. 1999, 285–296.