中国科学: 地球科学 2023 年 第 53 卷 第 11 期: 2591 ~ 2608
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文

基于扩展卡尔曼滤波和自动微分技术对陆面数据同化系统参数的快速估计及其影响

田佳鑫[1], 卢麾[1,5*], 阳坤[1,2], 秦军[3], 赵龙[4], 周建宏[1], 姜兆志[1], 马小刚[1]

1. 清华大学地球系统科学系, 地球系统数值模拟教育部重点实验室, 清华大学全球变化研究院, 北京 100084;
2. 中国科学院青藏高原研究所, 青藏高原地球系统与资源环境国家重点实验室, 国家青藏高原科学数据中心, 北京 100101;
3. 中国科学院地理科学与资源研究所, 资源与环境信息系统国家重点实验室, 北京 100101;
4. 西南大学地理科学学院, 重庆金佛山喀斯特生态系统国家野外科学观测研究站, 重庆 400715;
5. 清华大学-西安市勘察测绘院新型智能测绘联合研究中心, 北京 100084
* 通讯作者, E-mail: luhui@tsinghua.edu.cn

收稿日期: 2022-11-20; 收修改稿日期: 2023-07-27; 接受日期: 2023-08-29; 网络版发表日期: 2023-10-25
第二次青藏高原综合科学考察研究项目(编号: 2019QZKK0206)、国家重点研发计划项目(编号: 2022YFC3002901)、国家自然科学基金项目(批准号: 42271491)和中国科学院国际合作项目(编号: 183311KYSB20200015)资助

摘要 土壤水分在干旱监测、洪水预报和水资源管理中发挥着重要作用. 数据同化方法能结合模型和遥感的优势, 生产高精度、时空连续的土壤水分数据. 目前, 陆面数据同化面临的主要难题之一是模型和观测误差的准确估计. 为此, 我们前面提出了双循环同化算法, 实现了模型和观测误差、模型算子和观测算子参数的同时估计. 由于该算法需要设置较大的集合个数才能实现稳定的参数估计, 导致参数估计效率低, 限制了其在大范围陆面同化系统中的应用. 为此, 本研究通过: (1) 采用二元数自动微分方法计算陆面过程模型(LSM)的雅可比矩阵以替代获取复杂的LSM的切线性模型; (2) 采用扩展卡尔曼滤波(EKF)框架替代集合卡尔曼滤波框架, 提高了双循环算法的参数估计效率. 在模型积分步长为1h和参数优化窗口为90天的条件下, 改进后的双循环算法参数优化效率提高近60倍. 为检验双循环算法在区域尺度陆面数据同化中的效果, 我们在青藏高原区域同化SMAP土壤水分产品, 并利用四个不同地理气候环境下的土壤水分观测网数据验证同化结果. 结果显示, 在青藏高原区域, 相较于基于默认参数的模型模拟结果, 基于EKF改进的双循环同化算法能够校正模型参数和观测的偏差, 估计的土壤水分精度更高, 同时也提高了地表温度和蒸散发的估计精度. 此外, 相较于全球土壤数据中的沙土含量、黏土含量以及孔隙度, 该方案得到的土壤参数空间分布更为合理. 本研究改进的双循环算法有望进一步提高陆面、生态以及水文数据同化技巧.

关键词 土壤水分, 数据同化, 参数校正, 偏差校正, 误差估计, 自动微分

中文引用格式: 田佳鑫, 卢麾, 阳坤, 秦军, 赵龙, 周建宏, 姜兆志, 马小刚. 2023. 基于扩展卡尔曼滤波和自动微分技术对陆面数据同化系统参数的快速估计及其影响. 中国科学: 地球科学, 53(11): 2591–2608, doi: 10.1360/SSTe-2022-0372
英文引用格式: Tian J, Lu H, Yang K, Qin J, Zhao L, Zhou J, Jiang Y, Ma X. 2023. Quick estimation of parameters for the land surface data assimilation system and its influence based on the extended Kalman filter and automatic differentiation. Science China Earth Sciences, 66(11): 2546–2562, https://doi.org/10.1007/s11430-022-1180-8

© 2023 《中国科学》杂志社

1 引言

土壤水分在全球水、能量和碳循环中扮演着重要角色, 控制着降水在地表的产流和下渗过程, 影响土壤蒸发和植被蒸腾、光合作用、呼吸等过程(Dorigo等, 2017). 鉴于土壤水分在陆气相互作用过程中的重要作用, 它被视为重要的气候变量之一(Wagner等, 2012).

基于遥感观测和陆面过程模式(land surface models, LSM)模拟是目前获取大范围土壤水分数据的主要方式, 但是皆未臻完美. 数据同化技术可以克服它们各自在估算土壤水分方面的局限性, 实现模型和观测数据的有机融合, 进一步提高土壤水分估计精度(李新等, 2020, 2022; Gettelman等, 2022). 陆面数据同化系统(land data assimilation system, LDAS)能将不同来源、不同分辨率的观测和LSM进行最优融合, 产生时空连续、精度更高的地表状态变量(土壤水分、土壤温度、地表温度、雪水当量和树冠储水量等)和水热通量数据(显热通量、潜热通量、净辐射等)(Lu等, 2015; Zhao和Yang, 2018; Lu等, 2020; Seo等, 2021; Amazirh等, 2022).

遥感观测资料、LSM、同化算法是LDAS的三大主要构成部分, LDAS的发展主要围绕这三部分进行. 近年来, 已成功发射的全球土壤水分探测卫星有土壤水分和海洋盐度卫星(SMOS; Kerr等, 2010)和主动被动土壤水分监测卫星(SMAP; Entekhabi等, 2010). 基于这些卫星观测和不同的反演算法, 各国已生产多套全球土壤水分产品, 如SMOS(SML2UDP; Kerr等, 2012)、SMOS-IC(Fernandez-Moran等, 2017)、SMAP-36km(SPL3SMP; O'Neill等, 2020)和SMAP-9km(SPL3SMP_E; Chan等, 2018). LSM是气候模式和数值天气预报模式的陆面分量, 能够进行精细描述陆地表面的物理、化学和生物过程(Liang等, 1996; Oleson等, 2010; Niu等, 2011). 同化算法控制着观测信息对模型状态变量的校正程度, LDAS用到的同化算法也从早期较为简单的最优插值法和牛顿松弛法升级为更加高级的同化算法, 例如扩展卡尔曼滤波(EKF; Gelb, 1974)、粒子滤波(PF; Gordon等, 1993)、集合卡尔曼滤波(EnKF; Evensen, 1994)、三维变分(3D-Var; Courtier等, 1994)以及四维变分(4D-Var; Courtier等, 1998)等. 此外, 地球大数据的爆炸式增长, 机器学习和数据同化算法结合的数据融合方法也成为陆面数据同化研究的热点(Li等, 2023). 尽管如此, 陆面数据同化要实现模型-观测的最佳融合, 以及进一步提升LDAS的性能, 仍然面临一些难题: LSM参数不准确, 模型误差和观测误差难以确定, 卫星遥感反演值存在系统性偏差.

LSM参数不准确是陆面过程模拟不确定性的主要来源之一. 以土壤参数为例, 土壤的热力和水力参数控制着能量和水分的传输, 不准确的土壤参数值可能会造成模拟结果存在偏差(Reichle等, 2004; Yang等, 2016). 在过去的几十年, 已有大量研究致力于开发高精度的全球土壤数据库, 例如, The Food and Agriculture Organization(FAO)发布了全球土壤类型和植被覆盖度数据(Batjes, 1997), 中山大学研究团队基于不同区域的土壤图和土壤剖面数据库研制了1km分辨率的全球土壤数据集(GSDE; Shangguan等, 2014; Dai等, 2019). 但是在某些区域, 这些数据中的土壤参数值仍与真实情况有较大差距. 为实现高精度的陆面过程模拟, 仍需要进行陆面过程模型参数校正. 模型参数校正方法主要有两种: 一是状态变量和模型参数联合估计方法, 该方法通过状态增广方法同时估计模型参数和状态变量(Moradkhani等, 2005; Xie和Zhang, 2013; Ait-El-Fquih等, 2016); 二是假设参数不随时间变化, 在优化窗口(年尺度)内通过最小化模拟值与观测值(站点观测或遥感反演值)之差来校正模型参数(Yang等, 2007; Zhao等, 2022).

模型预测和观测数据之间的最佳融合需要合理的模型误差和观测误差, 因为误差参数决定了在同化时刻是更多地相信模型还是观测(Orrell等, 2001; 李新等, 2021). 由于观测误差和模型误差来源复杂且随空间变化, 因此, 为LDAS输入合理的误差参数值是一项极其困难的任务(Dreano等, 2015). 为了估计模型误差, Dee(1995)提出了最大化新息似然函数的方法, 同时估计模型状态变量和模型误差, 但是该方法估计效率较低并且不能估计观测误差. Dreano等(2017)利用集合卡尔曼平滑(EnKS)和最大期望算法(EM), 实现了模型误差和观测误差的快速估计, 但是该算法缺少在实际陆面数据同化系统中的验证. 不同于以上方法, 三重组合方法利用三套独立的相同变量数据也能提供模型误差和观测误差参数, 被广泛应用于区域和全球尺度的数据同化研究中(Stoffelen, 1998; Chen等, 2018; Zhou等, 2021). 尽管如此, 同时获取相同变量的三套

独立数据的要求并不容易满足.

无偏误差是当前主流数据同化方法的一个关键假设, 即误差必须是严格随机的, 并且观测值和模型估计值的平均值一致(Reichle和Koster, 2004). 然而, 由于部分参数(如表面粗糙度参数、植被光学参数)的不确定性, 以及反演算法本身的缺陷, 当前被动微波反演的土壤水分产品在许多区域存在较大系统性偏差(Zhao等, 2021). 为了降低遥感资料的偏差对LDAS的影响, 需要对遥感土壤水分产品进行偏差校正. 在模型模拟无偏的前提下, 常用的偏差校正方法有累积分布函数匹配和线性缩放(Reichle和Koster, 2004; Draper等, 2011), 这两种方法通常在同化前用模型模拟值来校正观测的偏差, 但是不适用于模型和观测同时存在偏差的情景. 为此, Qin等(2022)假设模型模拟与卫星土壤水分产品存在线性关系, 提出了一种线性观测算子来同化土壤水分产品, 通过估计模型参数和观测算子参数, 以同时校正陆面过程模型和遥感土壤水分产品的偏差.

大量研究表明, 前人的工作主要针对误差估计、模型参数估计和观测偏差校正中的某一个问题开展. 然而, 在大尺度开展陆面数据同化时, 这些问题常常同时存在. Tian等(2022)发展的双循环算法能够在统一的同化框架内解决这三个问题. 但是, 该算法需要对EnKF设置较多的集合个数(N>400), 这导致参数估计效率过低. 为此, 本研究利用EKF算法替换EnKF, 并通过自动微分(AD)方法精确求解LSM的雅可比矩阵, 进而提高双循环同化算法的参数估计效率. 为了验证基于EKF的双循环算法, 我们在青藏高原上开展了SMAP土壤水分同化实验, 并根据站网观测和遥感地表参量产品, 综合评估LDAS对土壤水分、蒸散发、地表温度的估计精度.

2　基于EKF和AD改进双循环LDAS

双循环算法可以同时优化的静态参数(Θ)包括模型参数(ξ)、观测算子参数(ζ)、模型误差(Q)和观测误差(R). 从物理意义上来说, 观测算子参数、模型参数(孔隙度、土壤质地参数)在一定时间尺度内(年)保持恒定, 而模型误差和观测误差可能随时间变化. 为了实现误差参数和模型参数的同时优化, 本研究假设误差参数也是常数. 双循环LDAS参数优化过程如图1所示, 在一个框架内土壤水分同化和参数优化同时进行. 在

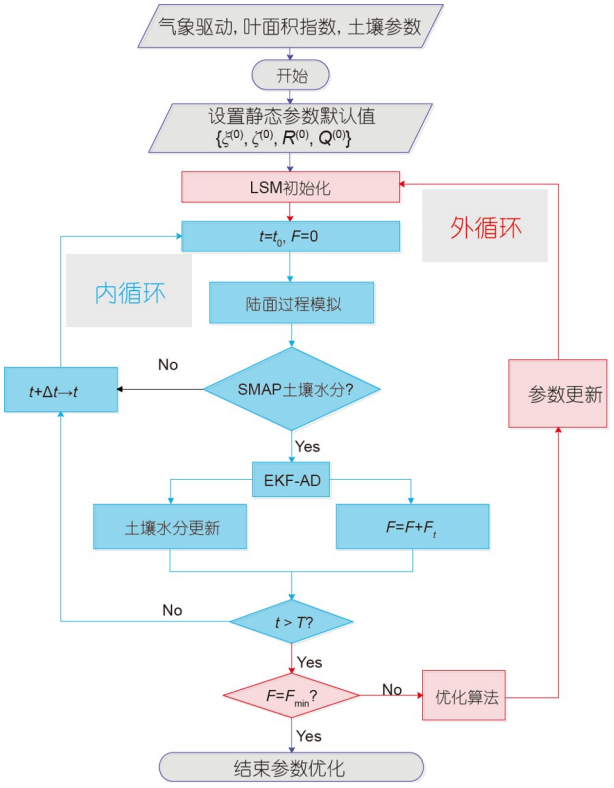

图 1　双循环陆面数据同化系统参数估计流程图

同化遥感观测期间, 内循环利用外循环传递的模型参数、观测算子参数、模型误差和观测误差参数, 当遥感观测可用时利用EKF更新LSM模拟的土壤水分, 并向外循环传递计算代价函数值; 内循环运行结束后, 外循环的优化算法模块判断代价函数是否达到最小, 若没有达到最小值, 则在参数空间寻找一组新的静态参数并传递给内循环, 内循环和外循环交替直至代价函数收敛(Tian等, 2022).

2.1　扩展卡尔曼滤波和自动微分

Evensen(1994)开创性地提出了EnKF, 以蒙特卡罗方法实现贝叶斯更新, 即利用集合方式描述模型先验和观测似然的概率分布, 然后通过贝叶斯定理得到后验估计值的概率分布. EnKF采用有限数目的集合预报结果来计算背景场误差协方差(P), 原理简单且易于实施, 已被广泛应用于陆面数据同化研究中. 然而, 在实际使用过程中为了降低采样误差对滤波器的影响, 通常需要设置较大的集合个数来保证EnKF的同化效果, 这增加了计算成本. 在EnKF之前, 非线性动力模型中

常用EKF实现卡尔曼滤波过程. EKF需要对非线性的模型算子和观测算子进行线性化, 其中的关键步骤是使用LSM的切线性模型(tangent linear model, TLM)计算LSM的雅可比矩阵. 然而, 开发和维护TLM过于复杂, 本研究通过自动微分算法求解LSM的雅可比矩阵.

2.1.1 非线性状态-空间系统的线性化

非线性系统的状态-空间方程可以表示为

$$x_k^f = M_k\left(x_{k-1}^a, u_k, \xi\right) + v_k, \tag{1}$$

$$y_k^o = H_k\left(x_k^f, \zeta\right) + w_k, \tag{2}$$

式中, M_k为非线性的模型算子, H_k为非线性的观测算子函数. 公式(1)描述模型状态从$k-1$时刻到k时刻的演进过程, x_k^f为状态变量后验估计值, x_{k-1}^a为状态变量的先验预测值, u_k为气象驱动数据, ξ为模型参数, v_k为均值为0协方差为Q的模型误差. 公式(2)是联系状态变量与观测的观测算子, 其中y_k^o为观测值, ζ为观测算子的参数, w_k为均值为0协方差为R的观测误差.

如果非线性模型是连续可导函数, 通过泰勒级数展开可获得非线性函数在x_k^f的一阶泰勒级数, 那么$M(x)$和$H(x)$可表达为

$$M(x) = M\left(x_k^f\right) + M'\left(x_k^f\right)\left(x - x_k^f\right), \tag{3}$$

$$H(x) = H\left(x_k^f\right) + H'\left(x_k^f\right)\left(x - x_k^f\right). \tag{4}$$

此时, $M(x)$为仅包含一阶导数项和常数项的线性方程. 其中$\mathbf{M}'$和$\mathbf{H}'$分别模型算子和观测算子在x_k^f处的偏导数:

$$\mathbf{M}' = \left.\frac{\partial M_k}{\partial x_k}\right|_{\left(x_k^f, u_k, \xi\right)}, \tag{5}$$

$$\mathbf{H}' = \left.\frac{\partial H_k}{\partial x_k}\right|_{\left(x_k^f, \zeta\right)}. \tag{6}$$

当状态变量为矢量时, H'和M'可以表示为偏微分方程组的雅可比矩阵:

$$\mathbf{J}_M = \begin{bmatrix} \frac{\delta M_1}{\delta x_1} & \frac{\delta M_1}{\delta x_2} & \cdots & \frac{\delta M_1}{\delta x_n} \\ \frac{\delta M_2}{\delta x_1} & \cdots & \cdots & \frac{\delta M_2}{\delta x_n} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{\delta M_n}{\delta x_1} & \frac{\delta M_n}{\delta x_2} & \cdots & \frac{\delta M_n}{\delta x_n} \end{bmatrix}, \tag{7}$$

$$\mathbf{J}_H = \begin{bmatrix} \frac{\delta H_1}{\delta x_1} & \frac{\delta H_1}{\delta x_2} & \cdots & \frac{\delta H_1}{\delta x_n} \\ \frac{\delta H_2}{\delta x_1} & \cdots & \cdots & \frac{\delta H_2}{\delta x_n} \\ \vdots & \cdots & \cdots & \vdots \\ \frac{\delta H_n}{\delta x_1} & \frac{\delta H_n}{\delta x_2} & \cdots & \frac{\delta H_n}{\delta x_n} \end{bmatrix}. \tag{8}$$

本研究利用二元数自动微分(DNAD)方法来快速、精确求解$\mathbf{J}_M$和$\mathbf{J}_H$.

2.1.2 二元数自动微分

DNAD利用二元数算术运算规则和编程语言的运算符重载功能, 可以实现向前自动微分, 比其他近似求解一阶导数算法更准确和高效(Yu等, 2013).

假设一变量x, 其二元数可以定义为

$$x = x + x'd, \tag{9}$$

式中, x为实数部分; x'为对偶部分; d并不代表任何具体的值, 且$d^n=0$ ($n>1$). 为方便表达, 公式(9)右边可写为$\langle x, x'd\rangle$.

基于DNAD求解函数的导数过程如下, 假设模型算子为

$$M(x_1, x_2) = x_1x_2 + 2x_1, \tag{10}$$

式中x_1、x_2为独立变量. 基于DNAD计算函数值$M(x_1, x_1)$和对应的偏导数$\partial M/\partial x_1$、$\partial M/\partial x_2$, 首先将x_1、x_2写成二元数的形式:

$$x_1 = \langle x_1, x_1'd_1\rangle, \tag{11}$$

$$x_2 = \langle x_2, x_2'd_2\rangle. \tag{12}$$

然后将公式(11)、(12)带入到公式(10)中, 因为$d_1d_2=0$, 则有

$$\begin{aligned} &\left(x_1 + x_1'd_1\right)\left(x_2 + x_2'd_2\right) + 2\left(x_1 + x_1'd_1\right) \\ &= x_1x_2 + x_1x_2'd_2 + x_2x_1'd_1 + x_1'x_2'd_1d_2 + 2x_1 + 2x_1'd_1 \\ &= x_1x_2 + x_1x_2'd_2 + x_2x_1'd_1 + 2x_1 + 2x_1'd_1. \end{aligned} \tag{13}$$

公式(13)可以整理成:

$$M(x_1, x_2) = x_1x_2 + 2x_1 + \left(x_1x_2'\right)d_2 + \left(x_2x_1' + 2x_1'\right)d_1. \tag{14}$$

在计算$\partial M/\partial x_1$时需设置x_1'等于1, x_2'等于0; 而在计算$\partial M/\partial x_2$时候设置x_1'等于0, x_2'等于1; 则有

$$\partial M/\partial x_1 = x_2 + 2, \tag{15}$$

$$\partial M/\partial x_2 = x_1. \tag{16}$$

$\partial M/\partial x_1$、$\partial M/\partial x_2$分别是d_1、d_2前面的系数，而$M(x_1,x_2)$的函数值为$x_1x_2+2x_1$.

2.1.3　扩展卡尔曼滤波(EKF)

EKF主要包括预测和更新两步，预测步骤由以下方程表示:

$$\mathbf{x}_k^f = \mathbf{M}(\mathbf{x}_{k-1}^a), \tag{17}$$

$$\mathbf{P}_k^f = \mathbf{M}'_k\mathbf{P}_{k-1}^a\mathbf{M}'^{\mathrm{T}}_k+\mathbf{Q}. \tag{18}$$

方程(17)模拟k时刻的状态变量，其中$\mathbf{x}$为的状态变量，上标f代表模型预测，a代表同化分析值. 方程(18)用于预测$k-1$时刻至k时刻状态协方差矩阵($\mathbf{P}$)，其中$\mathbf{M}'$是模型算子在x_k^f处的雅可比矩阵；$\mathbf{Q}$为模型误差协方差矩阵.

更新步骤通过以下两个方程完成:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k\left(H\left(\mathbf{x}_k^f\right)-\mathbf{y}_k^o\right), \tag{19}$$

$$\mathbf{P}_k^a = (1-\mathbf{K}_k\mathbf{H}'_k)\mathbf{P}_k^f. \tag{20}$$

公式(19)描述了用观测对状态变量校正的过程，其中$\mathbf{y}_k^o$是k时刻的观测值. 方程(20)用于状态误差协方差矩阵，$\mathbf{K}_k$是卡尔曼增益且由以下方程得到:

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}'^{\mathrm{T}}_k\left(\mathbf{R}+\mathbf{H}'_k\mathbf{P}_k^f\mathbf{H}'^{\mathrm{T}}_k\right)^{-1}, \tag{21}$$

式中，$\mathbf{R}$为观测误差协方差矩阵.

2.2　代价函数

双循环算法外循环使用的代价函数形式如下:

$$\begin{aligned} L(\boldsymbol{\Theta}) &\equiv p(\mathbf{y}_{1:K};\boldsymbol{\Theta}) \\ &= \prod_{k=1}^{K} p\left(\mathbf{y}_k \mid \mathbf{y}_{1:k-1};\boldsymbol{\Theta}\right), \end{aligned} \tag{22}$$

式中，$L(\Theta)$为参数的似然函数，$p\left(\mathbf{y}_t \mid \mathbf{y}_{1:k-1};\boldsymbol{\Theta}\right)$为当参数为$\boldsymbol{\Theta}$观测到$\mathbf{y}_k$的概率，$K$表示参数优化窗口. 当使用贝叶斯滤波同化算法时，公式(22)右侧为

$$\begin{aligned} p\left(\mathbf{y}_k \mid \mathbf{y}_{1:k-1};\boldsymbol{\Theta}\right) = &\frac{1}{(2\pi)^{\dim\boldsymbol{\varepsilon}_k/2}}\frac{1}{\sqrt{\left|\mathbf{H}'\mathbf{P}_k^f\mathbf{H}'^{\mathrm{T}}+\mathbf{R}\right|}} \\ &\times\exp\left[-\frac{1}{2}\boldsymbol{\varepsilon}_k^{\mathrm{T}}\left(\mathbf{H}'\mathbf{P}_k^f\mathbf{H}'^{\mathrm{T}}+\mathbf{R}\right)^{-1}\boldsymbol{\varepsilon}_k\right], \end{aligned} \tag{23}$$

式中$\boldsymbol{\varepsilon}$为观测的新息，上标T表示转置. 将公式(23)带入到公式(22)中，并且对公式两边求对数，可得到对数似然函数:

$$\begin{aligned} -\ell(\boldsymbol{\Theta}) = \sum_{k=1}^{K}\Bigg[&\frac{\dim\boldsymbol{\varepsilon}_k}{2}\log(2\pi)+\frac{1}{2}\log\left|\sum\nolimits_k\right| \\ &+\frac{1}{2}\boldsymbol{\varepsilon}_k^{\mathrm{T}}\left(\sum\nolimits_k\right)^{-1}\boldsymbol{\varepsilon}_k\Bigg], \end{aligned} \tag{24}$$

式中，$\sum_k = \mathbf{H}'\mathbf{P}_k^f\mathbf{H}'^{\mathrm{T}}+\mathbf{R}$. 本研究中求公式(24)的最小值采用的优化算法为Duan等(1994)发展的SCE_UA算法，该算法吸收了遗传算法的优点，能够解决高维参数的全局优化问题，且不需要对目标函数求偏导，是解决非线性约束最优化问题的有效方法.

2.3　模型算子

LSM是LDAS的核心部分，描述了地表物质能量交换过程，直接影响水热状态变量、通量的估计精度. 本研究使用的模型算子为简单生物圈模型(SiB2)，它是一个典型的“双源”模型，考虑了土壤和植被的热传导(Sellers等，1996). SiB2内包括三层土壤和林冠层，可以描述冠层的空气动力学、气孔导度-光合作用和辐射传输. 土壤水力和土壤热力参数控制土壤水和能量传输的关键参数，本研究中土壤水力参数通过Lawrence和Slater(2008)发展的土壤水力参数化方案计算；对于土壤热力参数，考虑到青藏高原地区的土壤中富含丰富的有机质，采用Chen等(2012)发展的考虑有机质的土壤热力参数化方案，其具体计算如下:

$$\rho_s c_s = \left[0.076+0.748\rho_d/\rho_w\right]\times10^6+4.195\times10^6\theta_{sfc}, \tag{25}$$

$$\lambda_s = \lambda_d+(\lambda_m-\lambda_d)\exp\left[0.36\left(1-1/\theta_{sfc}\right)\right], \tag{26}$$

$$\lambda_m = 0.5^{\theta_s}\left[(7.7^{\mathrm{sand}}2.0^{1-\mathrm{sand}})^{1-\mathrm{soc}}0.25^{\mathrm{soc}}\right]^{1-\theta_s}, \tag{27}$$

$$\lambda_d = (0.135\rho_d+64.7)/(2700-0.947\rho_d), \tag{28}$$

$$K_s = 7.0556\times10^{-6.884+1.53\times\mathrm{sand}}(1-\mathrm{soc})+10^{-4}\mathrm{soc}, \tag{29}$$

$$\psi_s = -0.01\times10^{1.88-1.31\times\mathrm{sand}}(1-\mathrm{soc})-0.0103\times\mathrm{soc}, \tag{30}$$

$$b = (2.91+0.159\times\mathrm{clay})(1-\mathrm{soc})+2.7\times\mathrm{soc}, \tag{31}$$

式中，$\rho_s c_s$(J K^{-1}m^{-3})为土壤热容，θ_{sfc}(cm^3 cm^{-3})为表层土壤水分，λ_s(W m^{-1} K^{-1})为土壤热传导率，λ_m和λ_d分别为土壤在干燥和饱和情况下的热传导率；θ_s是土壤孔

隙度, sand(%)是沙土含量, clay(%)是黏土含量, soc(%)为有机碳体积含量, K_s(m s^{-1})、ψ_s(m)、b是水力参数, ρ_d为干土密度; ρ_w(kg m^{-3})为水的密度; ρ_s (kg m^{-3})为土壤颗粒的密度.

Yang等(2016)通过减少待估计参数的个数减弱异参同效对参数估计过程的影响, 通过沙土含量和土壤孔隙度计算ρ_d和soc:

$$\rho_d = \rho_s(1-\theta_s), \tag{32}$$

$$\text{soc}=[\theta_s-(0489-0.126\times\text{sand})]/(0.411+0.126\times\text{sand}). \tag{33}$$

因为LMS的土壤水热参数高度依赖于(sand, clay, θ_s), 本研究中将在双循环算法中优化这三个参数, 然后通过方程(25)~(33)转换成LSM所需的土壤水力和热力参数.

2.4 观测算子

被动微波传感器的探测深度随下垫面发生变化, 这会导致卫星土壤水分产品代表的深度和LSM设置的表层土壤厚度不一致, 从而导致反演值和模拟值之间的代表性误差. 此外, 被动微波土壤水分产品在部分区域可能存在系统性偏差. 为此, Qin等(2022)假设土壤水分模拟值和反演值之间存在线性关系, 提出一种线性的观测算子来同化遥感土壤水分产品. 线性观测算子定义如下:

$$w_{s,k} = \alpha w_{l,k} + \beta + v_k, \tag{34}$$

式中, α和β为偏差校正系数, $w_{l,k}$是表层土壤水分模拟值, $w_{s,k}$是卫星土壤水分反演值, v_k是随机观测误差, 且v~$N(0, R)$.

3 研究区域与数据

3.1 土壤温湿度观测网

青藏高原是世界上平均海拔最高的高原. 近10年来, 为了支持被动微波遥感和陆面数据同化的发展, 国内外学者合作在高原上建立了多个土壤水分观测网(Ma等, 2003; Su等, 2011; Zhao等, 2013; Lan和Liu, 2018). 本研究搜集了4个观测时间相对较长、不同气候条件下观测网的观测数据用于验证同化结果. 图2显示了4个观测网的地理位置, 即青藏高原西部的狮泉河观测网(WTP-Shiquanhe)、青藏高原南部的帕里观测网(STP-Pali)、青藏高原中部的那曲观测网(CTP-Naqu)和青藏高原东部的玛曲观测网(ETP-Maqu), 它们分别代表了四种气候条件: 干旱气候、半干旱气候、半湿润半干旱气候和半湿润气候. 下面将具体介绍这四个监测网的仪器配置和相应的地理气候条件.

青藏高原西部的狮泉河土壤温湿度观测网(WTP-Shiquanhe)建立于2010年, 也是这四个观测网中最为干旱的网络. 狮泉河地区位于青藏高原西部, 年平均降水量约为100mm, 因此该地区下垫面的特点是草木稀少, 甚至是沙漠. 该网络内布置了20个土壤温湿度仪器, 仪器型号为EC-TM, 测量深度为5、10、20、40、60或80cm. 由于该网络位于偏远地区, 维护困难、年代久远, 导致部分土壤水分站点观测不连续. 因此, 本文选取12个在2018年6月8日数据完整的站点的观测资料. 关于该网络的更多细节可以参考Zhang等(2021).

青藏高原南部的帕里土壤温湿度观测网(STP-Pali)始建于2015年, 位于喜马拉雅山脉北部. 该地区地形最为复杂, 气候类型为半干旱气候, 下垫面为稀疏草地, 年平均降水量约为200~300mm. 该网络内布置了24个土壤温湿度站点, 分别测量0~5、10、20和40cm深度的土壤水分和土壤温度, 传感器类型为EC-TM, 记录频率为30min. 关于STP-Pali的详细介绍可以参考Chen等(2017).

青藏高原中部的那曲土壤温湿度观测网(CTP-Naqu)始建于2010年, 在100km×100km范围内布置了56个土壤温湿度站点. 该地区处于青藏高原中, 平均海拔约为4500m, 地形相对平坦, 主要为平缓的丘陵, 下垫面主要为高寒草甸, 年平均降水量约为400~500mm, 气候类型为半干旱半湿润气候. CTP-Naqu观测网的特点在于设置了三个尺度嵌套网络(1°×1°、0.3°×0.3°、0.1°×0.1°), 可以用于验证不同分辨率的气候模型和遥感反演产品. 该观测网络主要仪器是EC-TM和5TM(METER, Inc., USA), 测量深度分别为0~5、10、20和40cm, 数据记录频率为30min (Yang等, 2013). 本研究中使用构成大尺度网络(1°×1°)的35个站点的土壤水分观测作为验证数据.

青藏高原东部的玛曲土壤温湿度观测网(ETP-Maqu)于2008年建成, 覆盖面积为125km×75km. 玛曲地区位于青藏高原东部, 平均海拔大约为3500m, 年降水量约为600mm, 气候类型为半湿润气候, 下垫面为高寒草甸. 该网络内安装了24个站点, 仪器型号为EC-

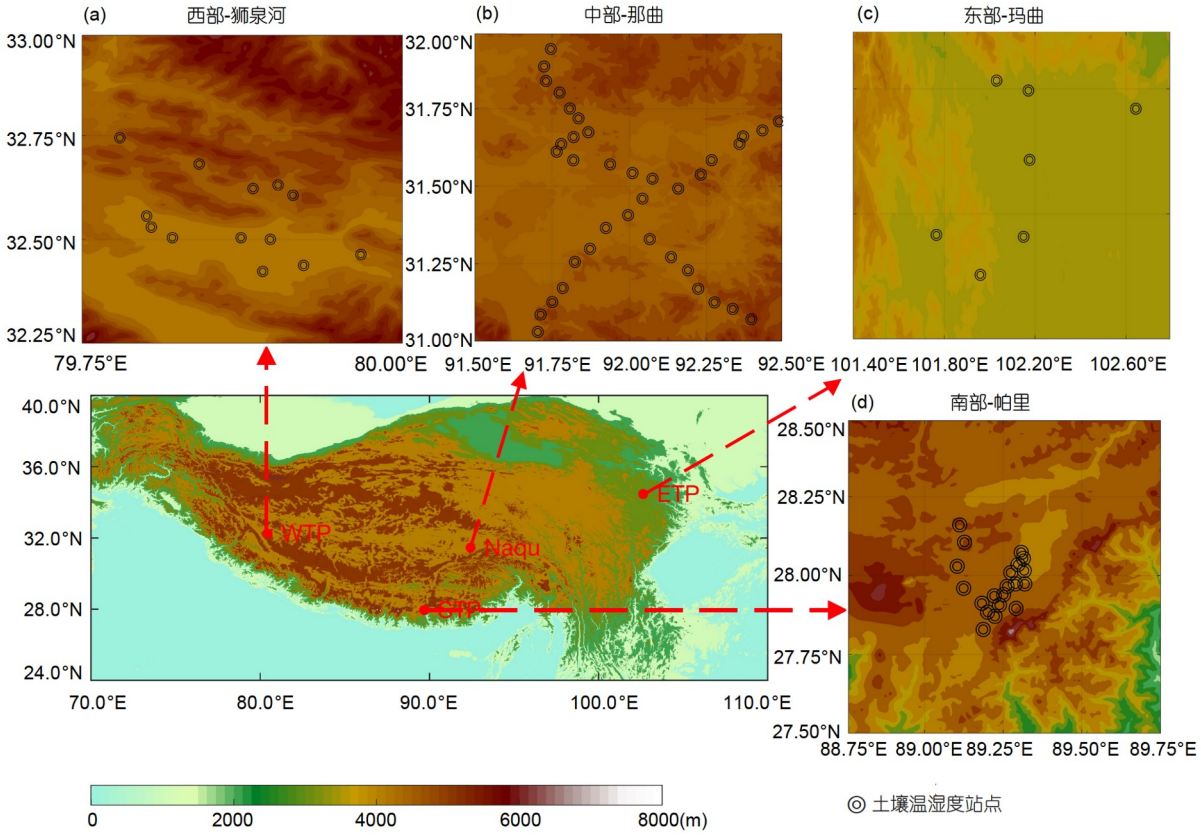

图 2　青藏高原地形与土壤温湿度观测网分布

(a) 西部-狮泉河(WTP-Shiquanhe); (b) 中部-那曲(CTP-Naqu); (c) 东部-玛曲(ETP-Maqu); (d) 南部-帕里(STP-Pali)

TM, 测量地表下5、10、20、40和80cm的土壤水分和土壤温度, 记录间隔为15min, 关于该观测网络的更多细节可以参考Su等(2011). 本文收集了在2018-06-08具有连续记录的7个站点的观测数据作为验证数据.

3.2　气象驱动数据

本研究采用中国区域高分辨率气象驱动数据(CMFD)和第三极地区长时间序列高分辨率降水数据(TPHiPr)共同驱动双循环LDAS. CMFD融合了卫星产品、再分析数据和站点观测, 包含7个气象变量, 分别为地表气压、2-m空气温度、空气湿度、10-m风速、降水、向下短波辐射和向下长波辐射. CMFD覆盖中国大陆, 时间跨度为1979~2018年, 时间分辨率为3h, 空间分辨率为0.1°×0.1°(He等, 2020). TPHiPr是以短期的高分辨率WRF模拟为基础, 利用机器学习方法对ERA5降水数据进行降尺度, 并融合了超9000个雨量站的观测数据, 得到的降水产品. TPHiPr覆盖第三极地区, 时间跨度为1979~2020年, 时间分辨率为1h, 空间分辨率是0.03°×0.03°(Jiang等, 2023). Gao等(2020)指出CMFD的降水数据在复杂地形地区存在不确定性, 而降水又是影响土壤水分的关键气象因子. 此外, 评估结果表明TPHiPr在第三极地区比目前主流的格点降水产品具有更高的精度. 为此, 本研究采用TPHiPr降水产品来替代CMFD的降水数据, 而其他气象变量保持不变.

3.3　遥感土壤水分、地表温度和蒸散发数据

本研究在双循环LDAS中同化SMAP被动微波土壤水分产品(SPL3SMP; O'Neill等, 2020). SMAP于2015年1月在美国戈达德航天中心发射升空, 用于监测全球土壤水分、土壤冻融和海洋盐度, 它同时搭载L波段辐射计和L波段的雷达, 轨道为太阳同步轨道, 重返周期为2~3天, 升轨为地方时6:00AM, 降轨为地方时6:00PM, 目标土壤水分反演精度无偏均方根误差

小于0.04cm^3 cm^{-3}(Entekhabi等, 2010).

本文用MODIS LST产品作为地表参考值评估LDAS估算的地表温度(LST). 因为基于昼夜算法生产MODIS 5km分辨率的LST比分裂窗算法反演的1km LST产品受到像元内云的影响更小. 本研究选用了MODIS/Aqua卫星生产的0.05°分辨率的LST产品(MYD11C1; Wan, 2007). Aqua可以提供白天和夜间的LST, 在太阳同步轨道的升轨通过赤道的时间约为地方时1:30PM, 降轨约为地方时1:30AM.

本研究利用PML_V2模型估算的蒸散发产品作为参考数据评估双循环LDAS的蒸散发. PML_V2使用一个水碳耦合冠层导度模型来估计蒸散发和总初级生产力, 以GLDAS 2.1的气象驱动和MODIS叶面积指数、反射率和发射率作为模型输入. Zhang等(2019)将PML-V2结果与全球95个通量塔的8天测量对比, 然后校正PML_V2在10种植被功能型下的模型参数, 最后根据MODIS MCD12Q2.006IGBP分类将参数移植至全球, 生产了全球蒸散发与总初级生产力数据集, 空间分辨率为500m, 时间分辨率为8天, 时间跨度为2002年7月至2019年8月.

3.4 土壤参数数据

中山大学陆面过程研究组通过融合世界土壤图和多个区域和国家级土壤数据库, 生产了全球土壤数据集(GSDE). 该数据提供全球网格化的土壤信息, 例如土壤颗粒大小分布、养分、沙土含量、黏土含量和有机碳含量等, 水平空间分辨率为30″×30″, 共有8个垂直分层(Shangguan等, 2014). 本研究采用了GSDE的表层沙土含量、黏土含量和孔隙度数据, 然后结合公式(25)~(33)计算LSM所需的土壤水力和热力参数. 此外, 本研究收集了Chen等(2012)在青藏高原的土壤样品测量值用于验证双循环算法估计的沙土含量、黏土含量、孔隙度.

4 实验设计与结果

4.1 点尺度同化和区域尺度实验设计

针对单点尺度, 本文在CTP-Naqu观测网区域进行站点尺度的同化实验, 并对区域内的驱动数据、遥感观测、站点观测进行空间平均. 鉴于SiB2的土壤冻融方案无法模拟未冻水含量, 本研究仅选用夏季作为研究时段, 实验时间设置为2018-06-01~2018-08-31, 模型积分步长为1 h. LDAS优化的静态参数包括沙土含量(sand%)、黏土含量(clay%)、孔隙度(θ_s)、观测误差(R)、模型误差(Q)和观测算子参数(α、β). 基于上述设置, 本研究进行了两次同化实验和一次模拟实验: 实验1使用EnKF版本的双循环LDAS同化SMAP土壤水分产品(DA_EnKF), 集合个数设置为400; 实验2使用EKF版本的双循环LDAS同化SMAP土壤水分产品(DA_EKF); 实验3为默认参数条件下的陆面过程模拟(OL). 以上实验均以串行方式运行, 计算平台的CPU为24核心的2.3GHz Intel Xeon, 内存为128GB.

针对区域尺度, 本研究在整个青藏高原区域开展了SMAP土壤水分同化实验. 首先, 把不同分辨率的遥感资料(MODIS、SMAP、PML_V2)、气象驱动(TPHiPr、CMFD)和模型参数(GSDE)等数据的分辨率统一重采样到0.25°×0.25°. 然后, 实验区域设置为63.875°~103.875°E、25.625°~41.375°N, 同化时间为2018-06-01~2018-09-30, 模型积分步长为1h, 参数校正个数与点尺度实验保持一致. 最后, 进行陆面过程模拟实验(OL_TP)和SMAP土壤水分同化实验(DA_TP).

4.2 点尺度试验结果

表1列出了在被优化参数的默认值、估计值和测量值. 结果表明, 沙土含量和孔隙度的估计值比默认值更加接近测量值, 黏土含量估计值比测量值显著偏高; 两组同化实验中的模型误差(Q)均小于观测误差(R), 这表明在同化过程中更相信模型估计值; 而两个实验观测算子系数(α和β)的估计值之间的差异, 由两个实验不同的土壤参数导致. 图3对比了DA_EnKF和DA_EKF的参数优化时间, 发现后者的参数估计时间远低于前者. 图4对比了土壤水分估计值和站点观测的时间序列和散点图, 图4a中SMAP土壤水分反演值存在明显高估, OL土壤水分模拟值存在显著低估, 而DA_EKF和DA_EnKF的土壤水分估计值均较好地反映了CTP-Naqu区域真实的土壤水分变化. 图4b显示基于EKF的土壤水分估计值的R^2为0.84, RMSE为0.022cm^3 cm^{-3}, MBE为−0.0074cm^3 cm^{-3}; 图4c显示基于EnKF的土壤水分的R^2为0.86, RMSE为0.018cm^3 cm^{-3}, MBE为0.0048cm^3 cm^{-3}.

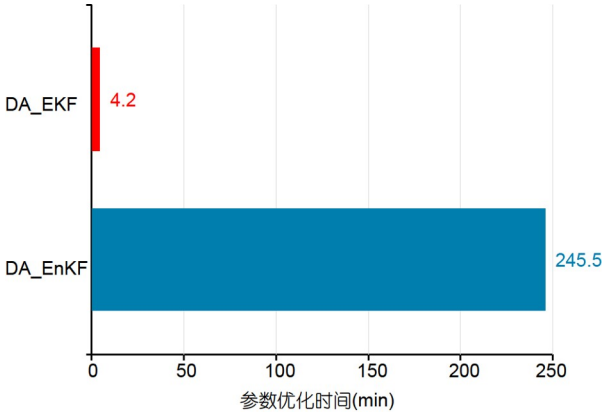

图 3　**基于EKF和EnKF的双循环LDAS参数优化时间对比**

4.3　区域尺度同化实验结果

图5展示了月平均SAMP土壤水分反演值、OL_TP模拟值和DA_TP估计值的空间分布. 可以看出, 土壤水分模拟值无法反映青藏高原夏季土壤水分东湿西干的空间分布特征; 而DA_TP估计值与SMAP反演值具有相似的空间分布特征, 但是两者土壤水分月均值仍然存在一定程度的差别. 在青藏高原中部和东部, DA_TP估计的土壤水含量高于SMAP反演值. 土壤水分估计值可能更准确地反映青藏高原中东部的土壤水分的空间变化, 因为青藏高原东部和东南部区降水丰沛且有大量的高寒草甸和森林下垫面; 被动微波亮温反演土壤水分会受植被影响, 在浓密植被下垫面土壤

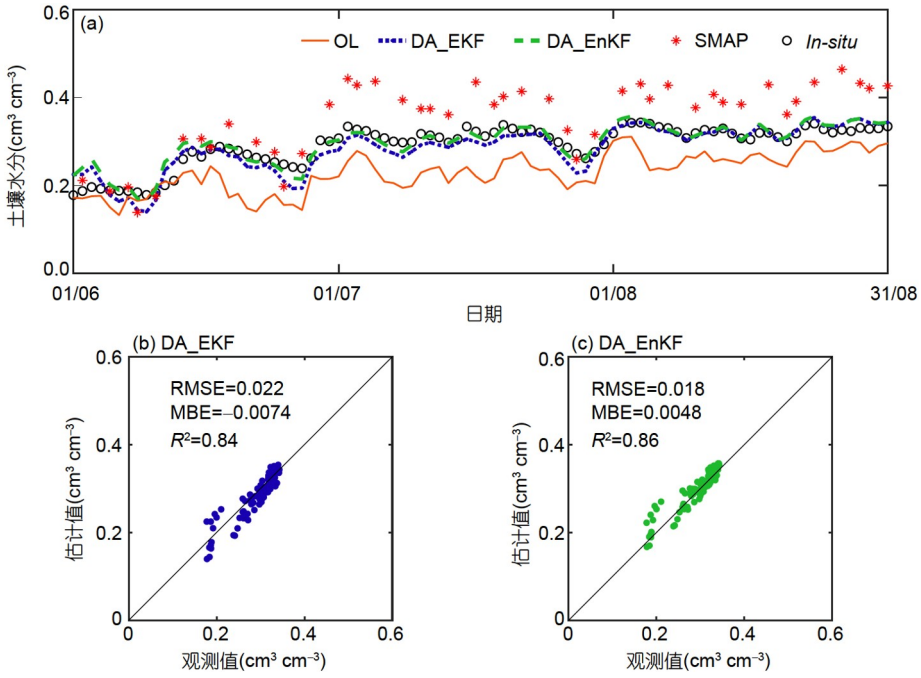

图 4　**土壤水分模拟值(OL)、同化实验估计值(DA_EKF、DA_EnKF)、SMAP反演值(SMAP)和站点平均实测(*in-situ*)的时间序列对比(a)以及DA_EKF估计值(b)和DA_EnKF估计值(c)与站点实测对比的散点图**

R^2为相关系数平方, RMSE为均方根误差, MBE为平均偏差

表 1　**模型参数、误差参数、观测算子参数的估计值和观测值对比**[a)]

实验	Sand	Clay	θ_s	Q	R	α	β
DA_EnKF	35.9	27.1	0.45	0.01	0.02	1.20	0.01
DA_EKF	41.82	14.11	0.42	0.008	0.03	1.42	−0.68
OL	38.89	11.33	0.38	—	—	—	—
测量值	69	6	0.51	—	—	—	—

a) 表中土壤质地参数和孔隙度参数的测量值来自Chen等(2012)

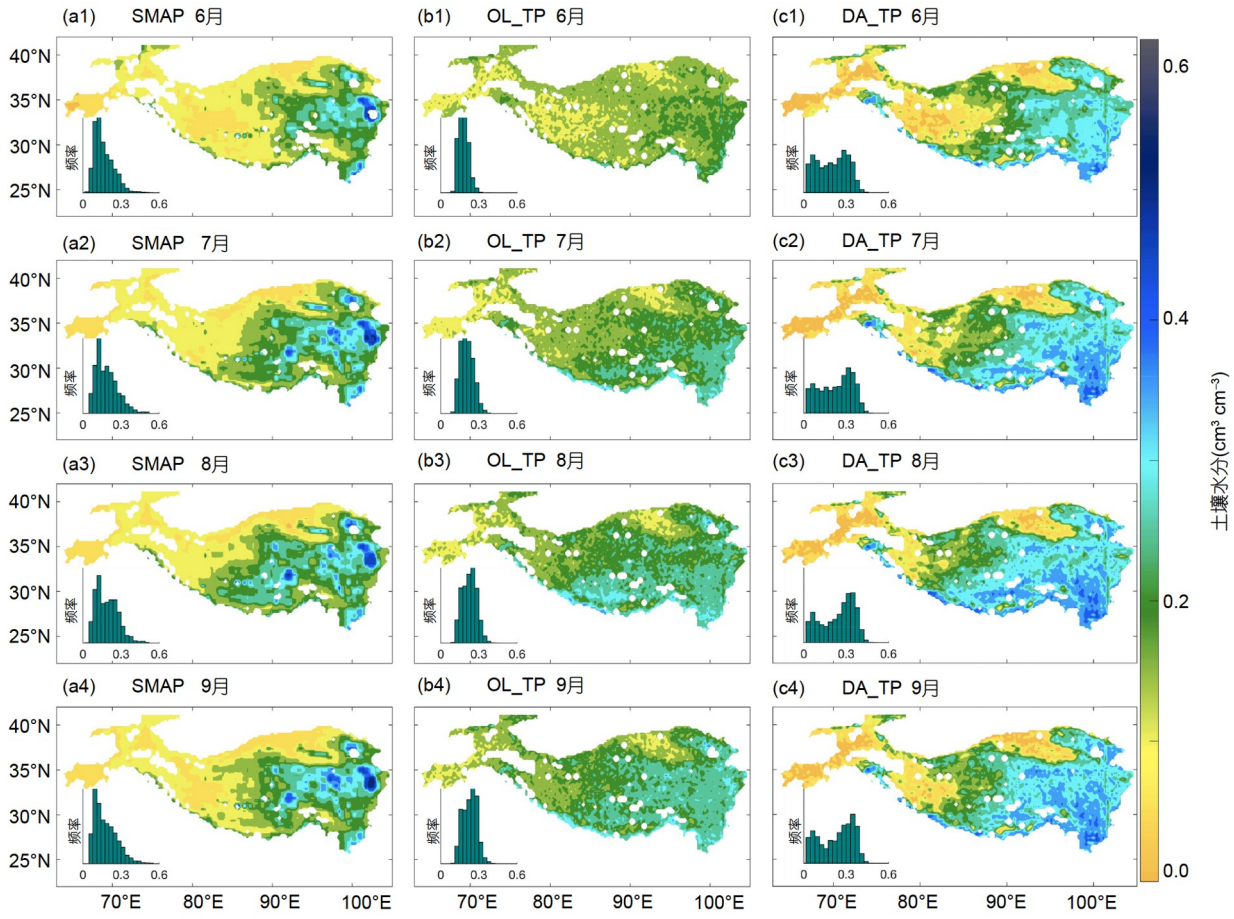

图 5　青藏高原土壤水分的空间分布

第一行至第四行分别代表6月、7月、8月、9月. (a1)~(a4) SMAP反演值; (b1)~(b4) OL_TP估计值; (c1)~(c4) DA_TP估计值

水分反演可能存在较大的不确定性(赵天杰, 2018). 此外, 卫星土壤水分产品的评估研究表明SMAP土壤水分产品在青藏高原存在显著的偏差, 并指出反演所使用的地表温度存在低估, 导致了SMAP土壤水分产品存在系统性偏低(Chen等, 2017; Zeng等, 2022). 以上对比分析不同于图4中所展示SMAP土壤水分存在高估, 因为土壤水分观测网仅仅可以获取小范围内相对准确的土壤水分数据, 在评估大尺度的遥感土壤水分产品时仍存在一定的空间局限性.

图6对比了土壤水分实测值、SMAP反演值、模拟值和同化估计值. 图6a1是它们在WTP-Shiquanhe区域的时间序列对比, 图6a2是估计值、模拟值和实测值的对比散点图; 图6b、6c和6d分别是在CTP-Naqu、ETP-Maqu和STP-Pali观测网的对比结果. 如图6a1所示, SMAP土壤水分产品在152~200天存在一定的高估, 但仍较好的反映出土壤水分的日变化; 而同化估计值没有完全消除模拟值的偏差, 这是因为驱动数据在该区域存在高估; 如图6a2所示, 相较于OL_TP, 土壤水分估计值的RMSE由0.075cm^3 cm^{-3}降低到0.034cm^3 cm^{-3}, MBE由0.072cm^3 cm^{-3}降低到0.016cm^3 cm^{-3}, R^2由0.56增加到0.83. 与WTP-Shiquanhe区域的结果类似, 在CTP-Naqu、ETP-Maqu、STP-Pali区域, 基于EKF的双循环LDAS可以有效提高土壤水分的估计精度, 并且有效地降低SMAP土壤水分偏差对同化结果的影响.

4.4　基于遥感资料评估双循环LDAS

为综合评估双循环LDAS同化土壤水分对其他状态变量或通量的影响, 本研究基于MODIS LST产品和PML_V2蒸散发产品评估了DA_TP实验和OL_TP实验

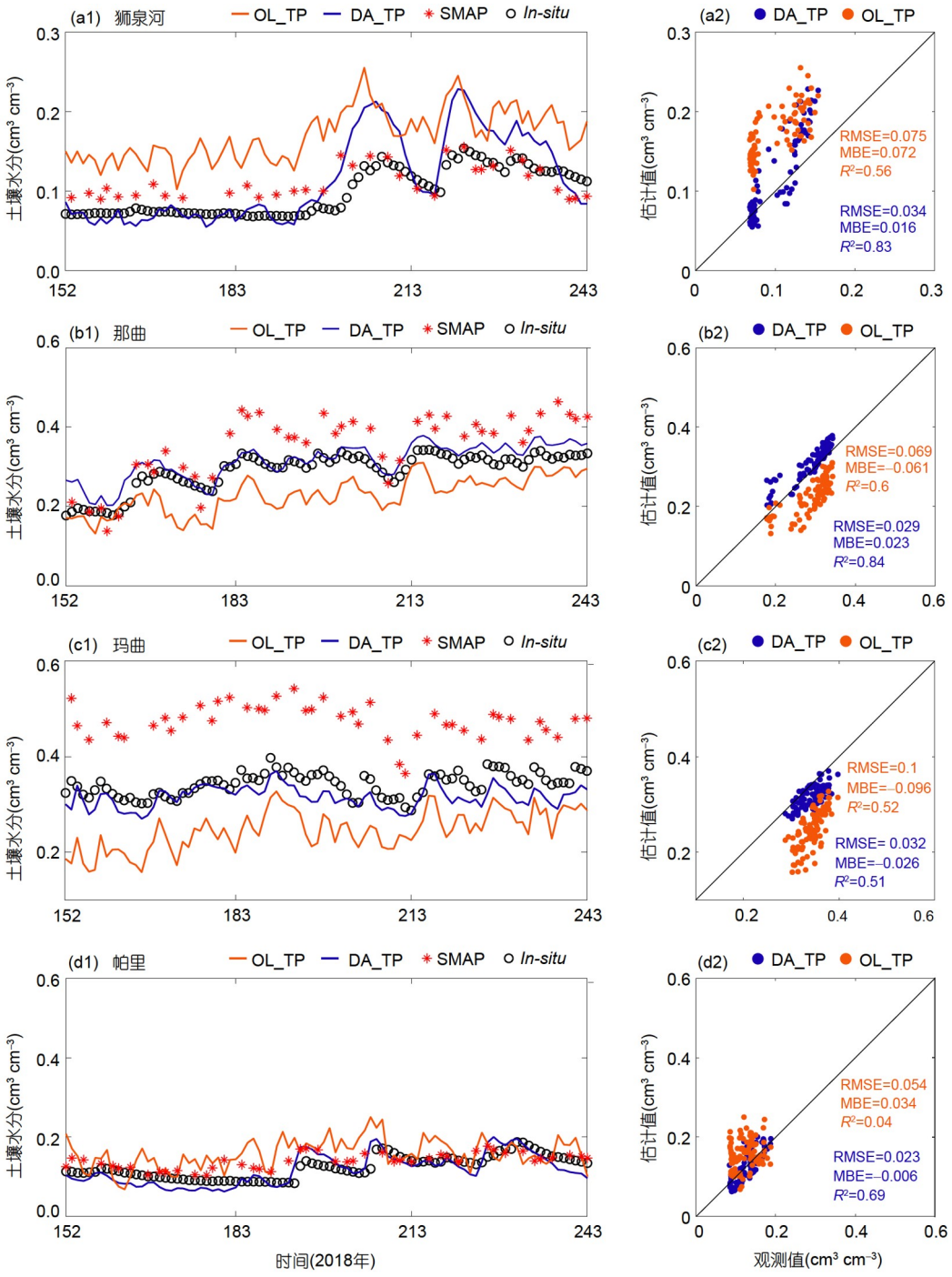

图 6　WTP-Shiquanhe(a)、CTP-Naqu(b)、ETP-Maqu(c)和STP-Pali(d)区域土壤水分模拟值、估计值、SMAP反演值和观测值的时间序列对比和散点图

生产的LST和蒸散发. 图7分别为DA_TP、OL_TP和PML_V2估算的累计蒸散发量空间分布. 结果显示, DA_TP实验与PML_V2的蒸散发产品空间分布特征相似, 而OL_TP实验估算的蒸散发无明显的空间分布特征, 例如干旱的塔里木盆地和高原西部的蒸散发量与高原中东部蒸散发相当. 图8展示了关于蒸散发与PLM_V2对比散点图, 结果显示, DA_TP的累计蒸散发量的空间相关系数远高于OL_TP, 并且DA_TP蒸散

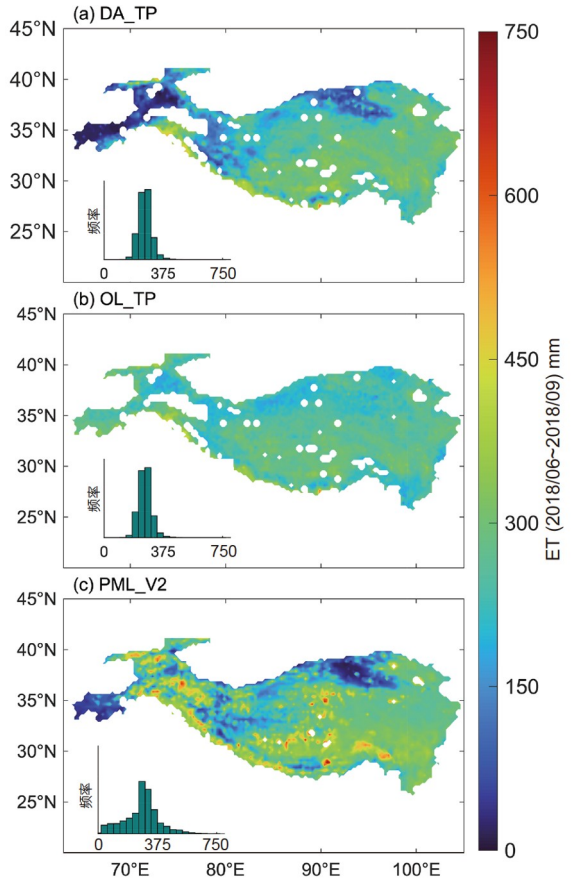

图 7　DA_TP(a)、OL_TP(b)和PML_V2(c)估算的青藏高原蒸散发空间分布

发的RMSE是106.66mm，而OL_TP的RMSE为113.34mm. 由于裸土蒸发和植被蒸腾过程会受到土壤水含量的影响，这些过程会影响地表物质能量的收支. 图9为2018年8月6日的LST模拟值和MODIS LST对比散点图，可以看出，DA_TP实验LST模拟值的RMSE在白天和夜间均小于OL_TP实验，并且白天LST模拟值的改进效果更显著，这是因为地表蒸散发主要发生在白天，而蒸散发会通过模型内的能量平衡和质量平衡影响LST的模拟.

4.5　静态参数估计值

图10展示了沙土含量、黏土含量和孔隙度的默认值和估计值的空间分布特征. 从图10a1和10b1可以看出，沙土含量的默认值和估计值具有类似的空间分布，均表现为在高原西部、北部沙土含量值较高，但是估计值整体上小于默认值; 沙土含量的均值由43.64%(图10a1)降低到40.45%(图10b1). 黏土含量的默认值大于估计值，均值从14.95%(图10a2)降低到12.4%(图10b2). 孔隙度是可能导致土壤水分模拟偏差的关键参数，对比图10a3和10b3发现，孔隙度默认值的均值为0.39，估计值的均值是0.44. 事实上，DA_TP估算的孔隙度空间分布更加合理，一是在四个土壤水分观测网的验证结果证明DA_TP实验的土壤水分估计值比OL_TP土壤水分模拟的精度更高，二是因为青藏高原东部的高寒区土壤中存有丰富有机碳，这会导致土壤孔隙度增加(Chen等, 2012).

图11为观测算子参数估计值的空间分布. 可以看出α代表的拟合系数在大部分高原区域小于1; 而β代表的截距量分布较随机. 如4.3节所述，基于线性观测

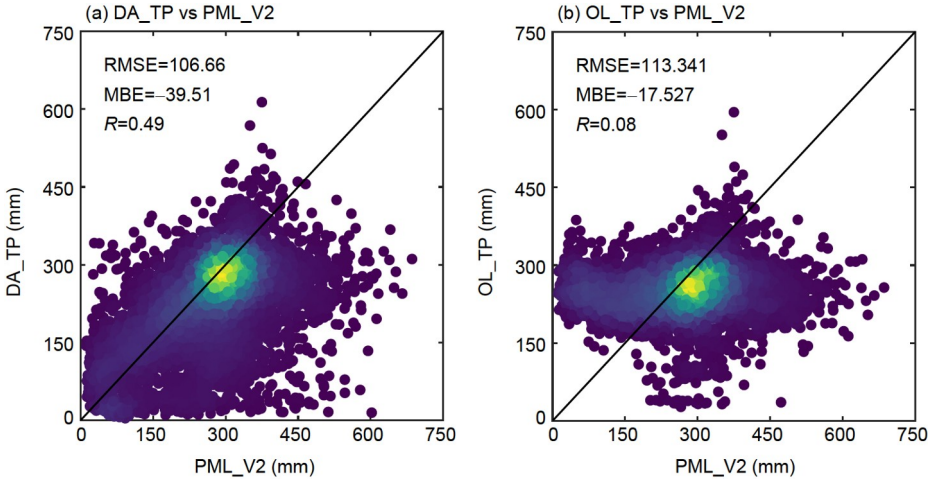

图 8　DA_TP(a)和OL_TP(b)实验生产的蒸散发与PML_V2蒸散发产品的对比散点图

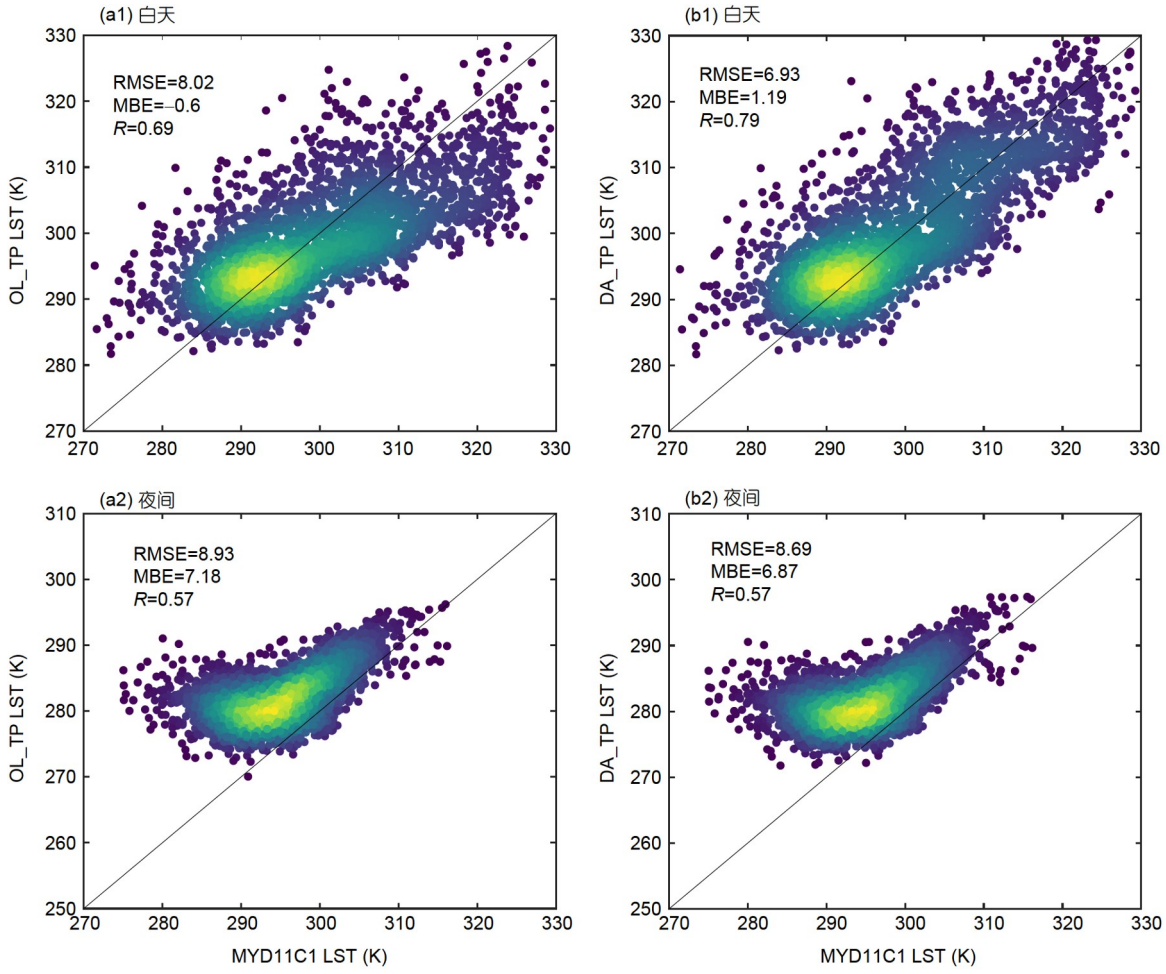

图 9　2018年8月6日 DA_TP和OL_TP的LST模拟值与相应MYD11C1日间((a1)和(b1))和夜间((a2)和(b2))LST产品的对比散点图

算子同化SMAP土壤水分，降低了卫星土壤水分偏差对同化系统的影响. 此外，基于优化的观测算子系数，可以在每个格点上校正SMAP土壤水分的偏差. 首先，根据观测算子求取它的反函数，然后将SMAP输入到反函数中，则可得到校正后的SMAP土壤水分($SMAP_{corr}$). 如图12所示，在CTP-Naqu和ETP-Maqu两个区域上SMAP、$SMAP_{corr}$和站点观测的土壤水分时间序列的对比图和散点图. 在CTP区域，SMAP土壤水分产品的MBE由0.062cm^3 cm^{-3}降低到0.024cm^3 cm^{-3}; 在ETP区域，SMAP的MBE从0.13cm^3 cm^{-3}降低到−0.017cm^3 cm^{-3}.

双循环算法还可以估计模型和观测的不确定性参数. 图13展示了模型误差(Q)和观测误差(R)在青藏高原上的空间分布. 显然，在DA_TP中的参数估计值在整个高原面上Q要比R更小，这表明在同化过程中EKF更加偏信模型.

5　结论

双循环LDAS在一个统一的框架内解决了当前陆面数据同化面临的三个棘手问题：误差参数估计、模型参数估计和观测偏差校正. 由于双循环算法原型采取了EnKF，需要开展较大的集合运算，参数估计效率过低，限制了其在区域同化中的应用. 为此，本研究利用EKF框架替换EnKF框架，从而提高双循环算法的参数估计效率；同时通过DNAD方法快速且精确地求解雅可比矩阵，避免开发LSM复杂的切线性模型. 为了检验双循环算法的改进效果，本研究首先在单点尺度

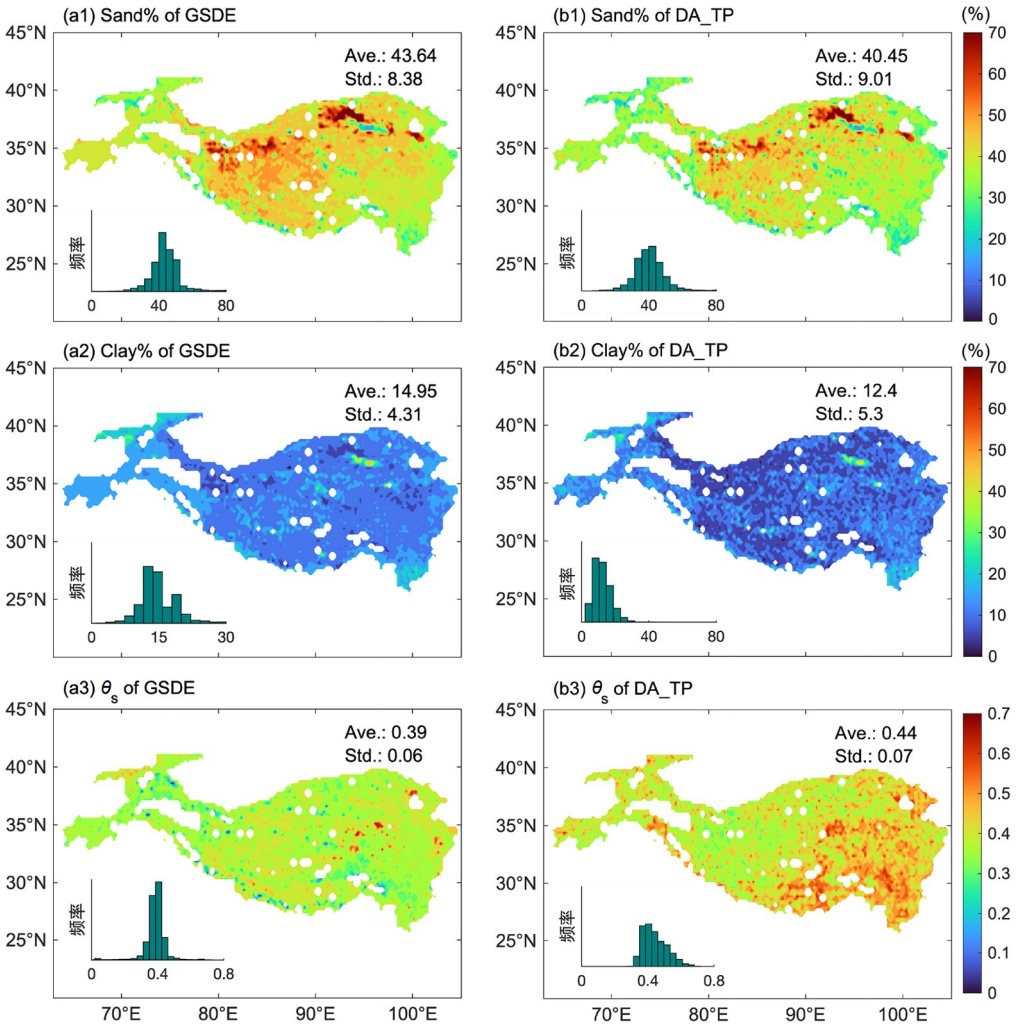

图 10　沙土含量((a1)和(b1))、黏土含量((a2)和(b2))以及孔隙度((a3)和(b3))默认值(左图)与估计值(右图)的空间分布

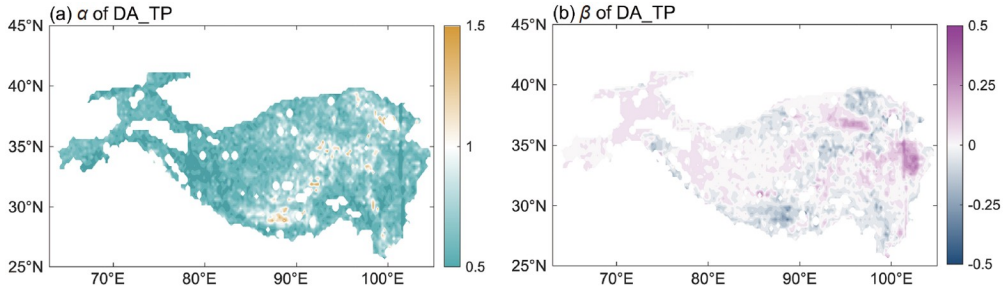

图 11　线性观测算子系数α(a)和β(b)估计值的空间分布

对比了基于EnKF和EKF-AD两个版本的双循环LDAS; 然后, 在整个青藏高原开展了SMAP土壤水分同化实验. 主要结论如下.

(1) 本研究通过DNAD快速和准确地求解观测算子和模型算子的雅可比矩阵, 避免开发复杂的切线性算子. 改进后的双循环算法在保证了土壤水分估计值

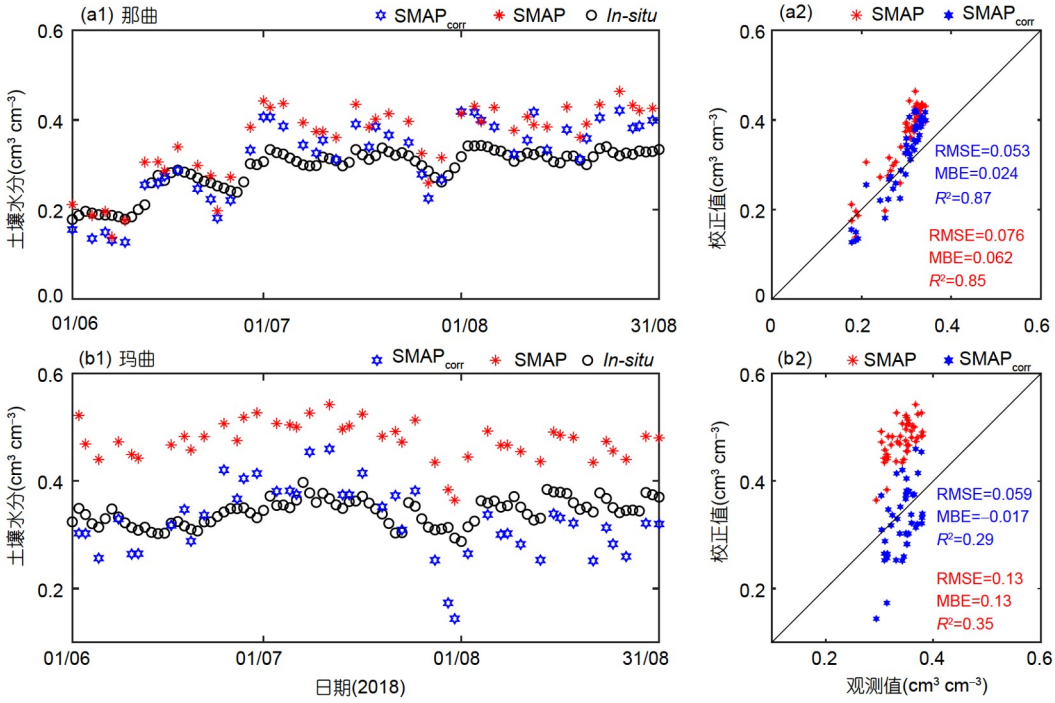

图 12　**CTP-Naqu(a)、ETP-Maqu(b)地区的SMAP、SMAP$_{corr}$与站点观测的土壤水分时间序列对比图和散点图**

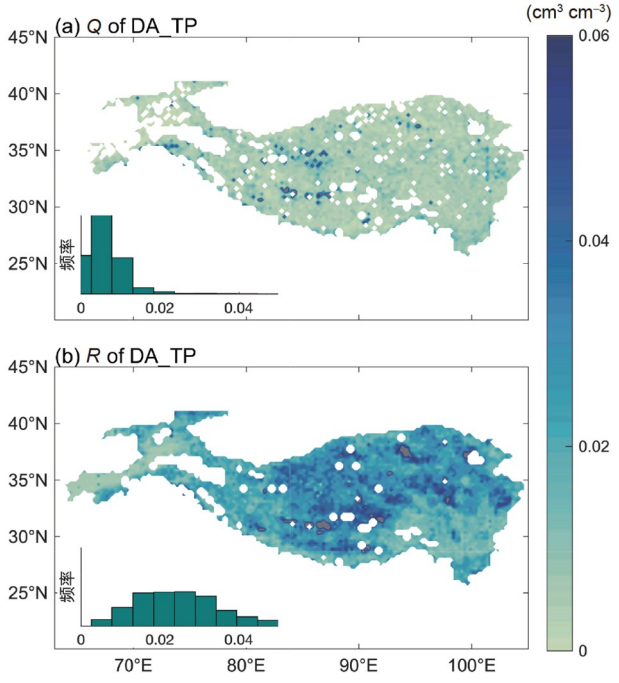

图 13　**模型误差(a)和观测误差(b)估计值的空间分布**

精度的前提下, 大幅提高参数估计效率. 在相同的参数优化时间窗口下，双循环同化算法参数优化效率提升近60倍，这为双循环LDAS的大尺度同化实验提供了必要条件.

(2) 利用青藏高原上四个不同气候类型的土壤水分观测资料验证双循环LDAS结果，验证结果显示相较于模拟实验，双循环算法有效地纠正了默认参数导致的模拟系统性偏差，显著地提高了土壤水分的估计精度. 此外，对LDAS的LST和蒸散发评估结果表明，同化SMAP土壤水分也提高了LDAS对LST和蒸散发的估计精度.

(3) 鉴于双循环算法能够同时估计模型误差、观测误差、模型参数以及观测算子参数，我们分析了各个参数估计值的空间分布状况，结果发现Q的估计值小于R，表明了在同化过程中LDAS更加偏信模型. 对比土壤参数的估计值和默认值发现，土壤孔隙度估计值在高原东部的空间分布比默认值更加合理. 此外，观测算子参数估计值可以用于校正SMAP的偏差.

综上，基于EKF-AD的双循环LDAS在青藏高原同化实验中展现了良好的同化性能，这可以为水文、生态、陆面数据同化提供了新的思路. 此外，双循环算法能够估计多源观测的误差，这有助于在LDAS框架下最优融合不同来源、不同分辨率遥感观测信息，生

产高精度、高时空分辨率的地表状态变量和水热通量.

参考文献

李新, 刘丰, 方苗. 2020. 模型与观测的和弦: 地球系统科学中的数据同化. 中国科学: 地球科学, 50: 1185–1194

李新, 马瀚青, 冉有华, 王旭峰, 朱高峰, 刘丰, 何洪林, 张臻, 黄春林. 2021. 陆地碳循环模型-数据融合: 前沿与挑战. 中国科学: 地球科学, 51: 1650–1663

李新, 郑东海, 冯敏, 陈发虎. 2022. 信息地理学: 信息革命重塑地理学. 中国科学: 地球科学, 52: 370–373

赵天杰. 2018. 被动微波反演土壤水分的L波段新发展及未来展望. 地理科学进展, 37: 198–213

Ait-El-Fquih B, El Gharamti M, Hoteit I. 2016. A Bayesian consistent dual ensemble Kalman filter for state-parameter estimation in subsurface hydrology. Hydrol Earth Syst Sci, 20: 3289–3307

Amazirh A, Er-Raki S, Ojha N, Bouras E, Rivalland V, Merlin O, Chehbouni A. 2022. Assimilation of SMAP disaggregated soil moisture and Landsat land surface temperature to improve FAO-56 estimates of ET in semi-arid regions. Agric Water Manage, 260: 107290

Batjes N H. 1997. A world dataset of derived soil properties by FAO-UNESCO soil unit for global modelling. Soil Use Manage, 13: 9–16

Chan S K, Bindlish R, O'Neill P, Jackson T, Njoku E, Dunbar S, Chaubell J, Piepmeier J, Yueh S, Entekhabi D, Colliander A, Chen F, Cosh M H, Caldwell T, Walker J, Berg A, McNairn H, Thibeault M, Martínez-Fernández J, Uldall F, Seyfried M, Bosch D, Starks P, Holifield Collins C, Prueger J, van der Velde R, Asanuma J, Palecki M, Small E E, Zreda M, Calvet J, Crow W T, Kerr Y. 2018. Development and assessment of the SMAP enhanced passive soil moisture product. Remote Sens Environ, 204: 931–941

Chen F, Crow W T, Bindlish R, Colliander A, Burgin M S, Asanuma J, Aida K. 2018. Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. Remote Sens Environ, 214: 1–13

Chen Y, Yang K, Qin J, Cui Q, Lu H, La Z, Han M, Tang W. 2017. Evaluation of SMAP, SMOS, and AMSR2 soil moisture retrievals against observations from two networks on the Tibetan Plateau. J Geophys Res-Atmos, 122: 5780–5792

Chen Y Y, Yang K, Tang W J, Qin J, Zhao L. 2012. Parameterizing soil organic carbon's impacts on soil porosity and thermal parameters for Eastern Tibet grasslands. Sci China Earth Sci, 55: 1001–1011

Courtier P, Andersson E, Heckley W, Pailleux J, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F, Fisher M. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. Q J R Meteorol Soc, 124: 1783–1807

Courtier P, Thépaut J N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. Q J R Meteorol Soc, 120: 1367–1387

Dai Y, Xin Q, Wei N, Zhang Y, Shangguan W, Yuan H, Zhang S, Liu S, Lu X. 2019. A global high-resolution data set of soil hydraulic and thermal properties for land surface modeling. J Adv Model Earth Syst, 11: 2996–3023

Das N N, Entekhabi D, Njoku E G. 2011. An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. IEEE Trans Geosci Remote Sens, 49: 1504–1512

Dee D P. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. Mon Weather Rev, 123: 1128–1145

Dorigo W, Wagner W, Albergel C, Albrecht F, Balsamo G, Brocca L, Chung D, Ertl M, Forkel M, Gruber A, Haas E, Hamer P D, Hirschi M, Ikonen J, de Jeu R, Kidd R, Lahoz W, Liu Y Y, Miralles D, Mistelbauer T, Nicolai-Shaw N, Parinussa R, Pratola C, Reimer C, van der Schalie R, Seneviratne S I, Smolander T, Lecomte P. 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Remote Sens Environ, 203: 185–215

Draper C, Mahfouf J F, Calvet J C, Martin E, Wagner W. 2011. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. Hydrol Earth Syst Sci, 15: 3829–3841

Dreano D, Mallick B, Hoteit I. 2015. Filtering remotely sensed chlorophyll concentrations in the Red Sea using a space–time covariance model and a Kalman filter. Spatial Stat, 13: 1–20

Dreano D, Tandeo P, Pulido M, Ait-El-Fquih B, Chonavel T, Hoteit I. 2017. Estimating model-error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximization algorithm. Q J R Meteorol Soc, 143: 1877–1885

Duan Q, Sorooshian S, Gupta V K. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J Hydrol, 158: 265–284

Entekhabi D, Njoku E G, O'Neill P E, Kellogg K H, Crow W T, Edelstein W N, Entin J K, Goodman S D, Jackson T J, Johnson J, Kimball J, Piepmeier J R, Koster R D, Martin N, McDonald K C, Moghaddam M, Moran S, Reichle R, Shi J C, Spencer M W, Thurman S W, Tsang L, Van Zyl J. 2010. The Soil Moisture Active Passive (SMAP) Mission. Proc IEEE, 98: 704–716

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res, 99: 10143–10162

Fernandez-Moran R, Al-Yaari A, Mialon A, Mahmoodi A, Al Bitar A, De Lannoy G, Rodriguez-Fernandez N, Lopez-Baeza E, Kerr Y,

Wigneron J P. 2017. SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product. Remote Sens, 9: 457

Gao Y, Chen F, Jiang Y. 2020. Evaluation of a convection-permitting modeling of precipitation over the Tibetan Plateau and its influences on the simulation of snow-cover fraction. J Hydrometeorol, 21: 1531–1548

Gelb A. 1974. Applied Optimal Estimation. Cambridge: MIT Press

Gettelman A, Geer A J, Forbes R M, Carmichael G R, Feingold G, Posselt D J, Stephens G L, van den Heever S C, Varble A C, Zuidema P. 2022. The future of Earth system prediction: Advances in model-data fusion. Sci Adv, 8: eabn3488

Gordon N J, Salmond D J, Smith A F. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc Radar Sonar Navigation, 140: 107–113

He J, Yang K, Tang W, Lu H, Qin J, Chen Y, Li X. 2020. The first high-resolution meteorological forcing dataset for land process studies over China. Sci Data, 7: 25

Kerr Y H, Waldteufel P, Richaume P, Wigneron J P, Ferrazzoli P, Mahmoodi A, Al Bitar A, Cabot F, Gruhier C, Juglea S E, Leroux D, Mialon A, Delwart S. 2012. The SMOS soil moisture retrieval algorithm. IEEE Trans Geosci Remote Sens, 50: 1384–1403

Kerr Y H, Waldteufel P, Wigneron J P, Delwart S, Cabot F, Boutin J, Escorihuela M J, Font J, Reul N, Gruhier C, Juglea S E, Drinkwater M R, Hahne A, Martín-Neira M, Mecklenburg S. 2010. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. Proc IEEE, 98: 666–687

Lan C, Liu Z. 2018. Soil temperature and moisture observation data of the Yarlung Zangbo River Basin (2017). National Tibetan Plateau/Third Pole Environment Data Center, doi: 10.11888/Hydro.tpdc.270032

Lawrence D M, Slater A G. 2008. Incorporating organic soil into a global climate model. Clim Dyn, 30: 145–160

Li X, Feng M, Ran Y, Su Y, Liu F, Huang C, Shen H, Xiao Q, Su J, Yuan S, Guo H. 2023. Big Data in Earth system science and progress towards a digital twin. Nat Rev Earth Environ, 4: 319–332

Liang X, Wood E F, Lettenmaier D P. 1996. Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. Glob Planet Change, 13: 195–206

Lu H, Yang K, Koike T, Zhao L, Qin J. 2015. An improvement of the radiative transfer model component of a land data assimilation system and its validation on different land characteristics. Remote Sens, 7: 6358–6379

Lu Y, Steele-Dunne S C, De Lannoy G J M. 2020. Improving soil moisture and surface turbulent heat flux estimates by assimilation of SMAP brightness temperatures or soil moisture retrievals and GOES land surface temperature retrievals. J Hydrometeorol, 21: 183–203

Ma Y, Su Z, Koike T, Yao T, Ishikawa H, Ueno K, Menenti M. 2003. On measuring and remote sensing surface energy partitioning over the Tibetan Plateau—From GAME/Tibet to CAMP/Tibet. Phys Chem Earth Parts A B C, 28: 63–74

Moradkhani H, Sorooshian S, Gupta H V, Houser P R. 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resources, 28: 135–147

Niu G Y, Yang Z L, Mitchell K E, Chen F, Ek M B, Barlage M, Kumar A, Manning K, Niyogi D, Rosero E, Tewari M, Xia Y. 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J Geophys Res, 116: D12109

Oleson K W, Lawrence D M, Gordon B, Flanner M G, Kluzek E, Peter J, Levis S, Swenson S C, Thornton E, Feddema J. 2010. Technical description of version 4.0 of the Community Land Model (CLM)

Orrell D, Smith L, Barkmeijer J, Palmer T N. 2001. Model error in weather forecasting. Nonlin Processes Geophys, 8: 357–371

O'Neill P, Chan S, Bindlish R, Chaubell M, Colliander A, Chen F, Dunbar S, Jackson T, Peng J, Mousavi M, et al. 2020. Soil Moisture Active Passive (SMAP) Project: Calibration and Validation for the L2/3_SM_P Version 7 and L2/3_SM_P_E Version 4 Data Products. Technical Report JPL D-56297. 1–44

Qin J, Tian J, Yang K, Lu H, Li X, Yao L, Shi J. 2022. Bias correction of satellite soil moisture through data assimilation. J Hydrol, 610: 127947

Reichle R H, Koster R D. 2004. Bias reduction in short records of satellite soil moisture. Geophys Res Lett, 31: L19501

Reichle R H, Koster R D, Dong J, Berg A A. 2004. Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. J Hydrometeorol, 5: 430–442

Sellers P J, Randall D A, Collatz G J, Berry J A, Field C B, Dazlich D A, Zhang C, Collelo G D, Bounoua L. 1996. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. J Clim, 9: 676–705

Seo E, Lee M I, Reichle R H. 2021. Assimilation of SMAP and ASCAT soil moisture retrievals into the JULES land surface model using the Local Ensemble Transform Kalman Filter. Remote Sens Environ, 253: 112222

Shangguan W, Dai Y, Duan Q, Liu B, Yuan H. 2014. A global soil data set for earth system modeling. J Adv Model Earth Syst, 6: 249–263

Stoffelen A. 1998. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. J Geophys Res, 103: 7755–7766

Su Z, Wen J, Dente L, van der Velde R, Wang L, Ma Y, Yang K, Hu Z. 2011. The Tibetan Plateau observatory of plateau scale soil moisture

and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products. Hydrol Earth Syst Sci, 15: 2303–2316

Tian J, Qin J, Yang K, Zhao L, Chen Y, Lu H, Li X, Shi J. 2022. Improving surface soil moisture retrievals through a novel assimilation algorithm to estimate both model and observation errors. Remote Sens Environ, 269: 112802

Wagner W, Dorigo W, de Jeu R, Fernandez D, Benveniste J, Haas E, Ertl M. 2012. Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci, I-7: 315–321

Wan Z. 2007. Collection-5 MODIS land surface temperature products users' guide. ICESS, University of California, Santa Barbara

Xie X, Zhang D. 2013. A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter. Water Resour Res, 49: 7350–7365

Jiang Y, Yang K, Qi Y, Zhou X, He J, Lu H, Li X, Chen Y, Li X, Zhou B, Mamtimin A, Shao C, Ma X, Tian J, Zhou J. 2023. TPHiPr: A long-term (1979–2020) high-accuracy precipitation dataset (1/30°, daily) for the Third Pole region based on high-resolution atmospheric modeling and dense observations. Earth Syst Sci Data, 15: 621–638

Yang K, Qin J, Zhao L, Chen Y, Tang W, Han M, Lazhu M, Chen Z, Lv N, Ding B, Wu H, Lin C. 2013. A multiscale soil moisture and freeze-thaw monitoring network on the third pole. Bull Amer Meteorol Soc, 94: 1907–1916

Yang K, Watanabe T, Koike T, Li X, Fujii H, Tamagawa K, Ma Y, Ishikawa H. 2007. Auto-calibration system developed to assimilate AMSR-E data into a land surface model for estimating soil moisture and the surface energy budget. J Meteorol Soc Jpn, 85A: 229–242

Yang K, Zhu L, Chen Y, Zhao L, Qin J, Lu H, Tang W, Han M, Ding B, Fang N. 2016. Land surface model calibration through microwave data assimilation for improving soil moisture simulations. J Hydrol, 533: 266–276

Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, Abraham A L, Penterman J, Fischer R L, Voinnet O, Navarro L. 2013. Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. Proc Natl Acad Sci USA, 110: 2389–2394

Zeng J, Shi P, Chen K S, Ma H, Bi H, Cui C. 2022. Assessment and error analysis of satellite soil moisture products over the third pole. IEEE Trans Geosci Remote Sens, 60: 1–18

Zhang P, Zheng D, van der Velde R, Wen J, Zeng Y, Wang X, Wang Z, Chen J, Su Z. 2021. Status of the Tibetan Plateau observatory (Tibet-Obs) and a 10-year (2009–2019) surface soil moisture dataset. Earth Syst Sci Data, 13: 3075–3102

Zhang Y, Kong D, Gan R, Chiew F H S, McVicar T R, Zhang Q, Yang Y. 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sens Environ, 222: 165–182

Zhao L, Yang K, He J, Zheng H, Zheng D. 2022. Potential of mapping global soil texture type from SMAP soil moisture product: A pilot study. IEEE Trans Geosci Remote Sens, 60: 3119667

Zhao L, Yang K, Qin J, Chen Y, Tang W, Montzka C, Wu H, Lin C, Han M, Vereecken H. 2013. Spatiotemporal analysis of soil moisture observations within a Tibetan mesoscale area and its implication to regional soil moisture measurements. J Hydrol, 482: 92–104

Zhao L, Yang Z L. 2018. Multi-sensor land data assimilation: Toward a robust global soil moisture and snow estimation. Remote Sens Environ, 216: 13–27

Zhao T, Shi J, Entekhabi D, Jackson T J, Hu L, Peng Z, Yao P, Li S, Kang C S. 2021. Retrievals of soil moisture and vegetation optical depth using a multi-channel collaborative algorithm. Remote Sens Environ, 257: 112321

Zhou J, Crow W T, Wu Z, Dong J, He H, Feng H. 2021. A triple collocation-based 2D soil moisture merging methodology considering spatial and temporal non-stationary errors. Remote Sens Environ, 263: 112509

(责任编委: 李新)