

基于波谱数据库先验信息的地表参数同化反演方法^{*}

王东伟 王锦地^{**} 肖志强 张戈

北京师范大学地理学与遥感科学学院, 遥感科学国家重点实验室, 环境遥感与数字城市北京重点实验室, 北京 100875

摘要 我国典型地物波谱数据库的构建为地表参数反演提供了先验信息, 而时间序列遥感观测数据则能够提供反演目标参数的动态变化信息. 文中将描述植被冠层反射的遥感辐射传输模型与描述作物生长变化的过程模型结合, 提出了综合利用时间序列遥感观测数据和波谱库所提供的参数先验信息对农作物生长过程中的关键结构参数叶面积指数(LAI)进行反演的方法. 时间序列遥感观测数据能为目标参数LAI的反演提供更多信息, 而作物生长模型的引入则能够对反演过程中LAI的变化趋势进行约束. 实验结果表明, 利用地物波谱库中提供的参数先验信息结合时间序列遥感观测数据, 采用变分数据同化算法可以明显提高具有时间变化特征的地表目标参数反演精度.

关键词 数据同化 作物生长模型 叶面积指数 先验信息

利用遥感观测数据提取地表参数的变化信息, 是遥感的基本问题. 遥感机理模型描述了遥感观测数据与地表参数之间的相互关系, 利用遥感观测数据通过模型反演提取地表参数是定量遥感的主要方法. 李小文等^[1]曾对基于先验信息的遥感反演方法进行了实验验证, 但基于先验信息遥感反演方法的实现需要从大量地表参数观测数据中提取先验信息, 而我国典型地物波谱数据库^[2]的建立, 实现了地物波谱和地表参数的配套测量与数据积累, 为构造地表参数反演所需的先验信息提供了有效数据源^[3].

数据同化的目的是尽可能融合多数据源实现目标参数的提取, 在考虑时间尺度变化模型的参数反演方面具有明显的优势. 自从 Le Demit^[3] 1982年首次将伴随算子引入气象资料的变分分析以来, 变分数据同化方法的研究和应用已经取得了很大的发展和完善. 变分数据同化方法由于充分地利用了时间序列观测信息和过程模型信息, 因而在气象、海洋等领域得到广泛的发展和应用. 随着各种对地观测卫星的升空, 对

同一个地区的重复观测(同一传感器或不同传感器)次数将会增多, 这为应用数据同化技术融合时间序列遥感数据反演地表参数提供了有力支持.

基于遥感辐射传输模型反演地表植被参数的问题之一是如何综合利用时间序列遥感观测数据和过程模型参数先验信息, 并遵从作物生长参数随时间变化的连续性特征. 将时间序列遥感观测数据与作物生长模型结合提取LAI已经引起一些研究者的重视, 主要的方法分为滤波和参数优化^[4,5]. 滤波方法一般是用遥感观测数据在作物生长模型模拟的相应时刻对模型的状态变量进行滤波调整. 而参数优化方法则是采用整个时间序列遥感观测数据来对作物生长模型的参数进行优化, 并根据最优化参数来提取目标参数. 例如 Doraiswamy^[6]采用MODIS LAI产品与作物生长模型的模拟LAI值构建以产量预报为目标的代价函数进行优化的方法; 而 Jarlan^[7]则采用SPOT/VEGETATION NDVI产品与作物生长模型结合通过同化的方法评估草的地上总

2007-11-30 收稿, 2008-02-27 收修改稿

^{*} 国家自然科学基金(批准号: 40571107)、国家重点基础研究发展计划(批准号: 2007CB714407)、高等学校博士学科点专项科研基金、长江学者和创新团队发展计划资助项目、国际科技合作重点项目(批准号: 2004DFA06300)和国家统计遥感前沿技术研究与应用示范(2006AA120108)资助项目

^{**} 通信作者, E-mail: wangjd@bnu.edu.cn

生物量, 其精度误差达到 404 kg/ha.

本文采用变分同化方法反演 LAI, 重点研究在遥感辐射传输模型反演中引入作物生长过程模型和时间序列遥感观测数据的方法. 作物生长模型往往是一个复杂的模拟系统, 经常会涉及植被结构模拟、植被生理过程模拟、土壤水平衡模拟、土壤 N 平衡模拟、以及植被与周围环境之间的能量交换等过程. 为突出研究重点, 本文暂没有采用复杂的作物生长机理模型, 例如 CERES, WOFOST 等, 而是采用简化的作物生长模型—LOGISTIC 模型, 旨在为采用复杂的机理性作物生长模型和时间序列遥感观测进行地表植被的参数提取提供方法验证结果.

本文所使用的简化作物生长模型涉及冬小麦的单个生长季, 时间尺度以天为基础. 反演目标参数为冬小麦主要生长期的 LAI, 反演代价函数则融合了作物生长过程模型单季模拟结果和生长季内时间序列冠层反射率观测数据; 在反演过程中, 随着时间序列观测反射率数据的逐次引入, 我们采用前一时间段的反演结果来为融合了新观测数据的代价函数提供更新后的待优化参数先验信息, 实现了对参数先验信息的更新. 本文所使用的时间序列观测为地物波谱库中冬小麦在主要生长期内的冠层反射率观测数据. 与只采用遥感辐射传输模型进行单个时刻 LAI 反演的结果相比, 同化反演结果在时间连续性和精度上都有明显改进.

1 遥感数据同化策略

为了综合利用时间序列反射率观测数据与作物生长模型反演冬小麦在主要生长期内的 LAI, 可以构建代价函数形式如下:

$$\begin{cases} J_1(u) = \frac{1}{2}(u - u_B)^T B^{-1}(u - u_B) + \\ \frac{1}{2} \int_0^{\tau} (y_t - H(x_t))^T Q_t^{-1}(y_t - H(x_t)) dt & (1) \\ G(t, u, x_t) = 0 & (2) \end{cases}$$

式中 u 是待优化参数, u_B 是待优化参数的先验值, B 是待优化参数的先验误差协方差; y_t 是 t 时刻的冠层反射率观测值, $H(x_t)$ 是 t 时刻观测算子映射值, 其作用是将待反演目标参数从待反演目标参数空间映射到

对应的反射率空间, Q_t 是 t 时刻冠层反射率观测值与观测算子的误差协方差; $G(t, u, x_t)$ 是描述目标参数随时间变化规律的过程模型. 其中 x_t 在本文中就是 LAI. 由(1)和(2)式可见这是一个强约束问题, 约束项就是(2)式. 在这个代价函数中, 我们希望作为目标参数的 LAI 符合作物生长模型所表达的趋势, 也就是说 LAI 的取值来源于 $G(t, u, x_t)$. 求解这样一个约束代价函数的最小值, 可以直接将(2)式代入到(1)式中形成一个复合函数进行最优化求解, 这样做的原因是(2)式的输入是待优化参数 u , 而(2)式的输出在本文中是目标参数 LAI, 目标参数 LAI 进而是观测算子 $H(x_t)$ 的输入, 因而代价函数(1)式中的 u 就通过(2)式和 $H(x_t)$ 转换到了反射率空间. 求解这样一个代价函数最小值的方法有很多种, 例如最速下降法, 共轭梯度法等. 在本文中, 我们采用 Lagrange 乘子函数法, 因为根据 Lagrange 原理^[8], 符合约束项(2)式使(1)式最小的最优待优化参数 u 使 $J_1(u)$ 与 G 的梯度方向相同

$$\Delta J_1(u) = \int_0^{\tau} \Delta(\lambda(t)G(t, u, x_t)) dt \quad (3)$$

这里 $\lambda(t)$ 是 Lagrange 乘子函数, 同时可以将代价函数(1)式和约束方程(2)写成如下(4)式的形式

$$\begin{aligned} J(u) = & \frac{1}{2}(u - u_B)^T B^{-1}(u - u_B) + \\ & \frac{1}{2} \int_0^{\tau} (y_t - H(x_t))^T Q_t^{-1}(y_t - H(x_t)) dt + \\ & \int_0^{\tau} \lambda(t)G(t, u, x_t) dt \end{aligned} \quad (4)$$

代价函数(4)式与(1), (2)式是等价的, 这是因为代价函数(4)式的极值点处要求 $\Delta J(u) = 0$, $\Delta J(u) = 0$ 意味着(4)式的右手项的梯度为 0, 也就是

$$\begin{aligned} \Delta \left[\frac{1}{2}(u - u_B)^T B^{-1}(u - u_B) + \right. \\ \left. \frac{1}{2} \int_0^{\tau} (y_t - H(x_t))^T Q_t^{-1}(y_t - H(x_t)) dt \right] = \\ - \Delta \left[\int_0^{\tau} \lambda(t)G(t, u, x_t) dt \right] \end{aligned} \quad (5)$$

(5)式表达的含义就是(3)式，注意到(5)式中的右手项有一个负号，我们在(3)式中将这个负号合并到了 $\lambda(t)$ 中，因为在不同时间处， $\lambda(t)$ 实际是代价函数在此时间处梯度的一个比例因子，因而在(3)式中将其统一化到 $\lambda(t)$ 中。这样使(4)式最小的 u 就是代价函数(1)，(2)式的最优值。由此，求解约束极值问题转化为求解 Lagrange 乘子问题，也就是说代价函数(4)式与代价函数(1)，(2)式是等价的。

求解极值的方法是对(4)式求一阶变分，这里由于 $G(t, u, x_t)$ 的输出是 LAI ，所以将作物生长模型改写为 $x_t = LAI = M(t, u)$ ，这样代价函数(4)式就完全表达为待优化参数的形式，变分式如下

$$\delta J' = B^{-1}(u - u_B) \delta u + \int_0^{\tau} \frac{\partial H(x_t)}{\partial x_t} Q_t^{-1} (y_t - H(x_t)) \delta x_t dt + \int_0^{\tau} \delta \mathbf{a}(t)(x_t - M(t, u)) dt - \int_0^{\tau} \lambda(t) M'(t, u) \delta u dt + \int_0^{\tau} \lambda(t) \delta x_t dt \quad (6)$$

对(6)式各项进行合并，并根据变分原理可以得到

$$u = u_B + B \int_0^{\tau} \frac{\partial H(x_t)}{\partial x_t} Q_t^{-1} (y_t - H(x_t)) M'(t, u) dt \quad (7)$$

(7)式便是 u 的最优解形式。

数据同化就是通过求得上述代价函数的最小值来获得最优的待优化参数，在大气同化中，这是一个按时间段演进的过程，前一个时间段获得的最优值作为下一时间段待优化参数的先验值。我们采用同样的方式，利用前一时间段的最优参数值和后验误差协方差作为下次参数优化的先验信息。本文中观测算子 H 采用植被冠层辐射传输模型-SAIL模型^[9]， y_t 为 t 时刻冠层反射率观测值，过程模型 G 采用简化的作物生长模型-LOGISTIC模型^[10]，它是一个描述作物群体生长的模型，本文用其描述 LAI 与作物生长相对积温之间

的关系。

在这个强约束问题中，我们选择冬小麦在整个生育期内的 LAI 最大值 LAI_{MAX} 作为待优化参数，将其带入LOGISTIC模型可以获得目标参数 LAI ，进而将 LAI 作为观测算子SAIL模型的输入参数计算模拟冠层反射率，由此可以得到模拟反射率与实际测量反射率数据的差值；随着时间序列冠层反射率观测值的逐次引入和过程模型对目标参数 LAI 随时间变化连续性的控制，通过模拟反射率数据与实际观测反射率数据的差值就可以计算得到代价函数值。在同化反演的运行过程中，从输入给定时刻遥感观测反射率数据开始，求解(7)式可以得到当前时刻的最优 LAI_{MAX} ，逐次加入之后序列的反射率观测数据进行代价函数优化求解，直至所有的观测反射率被引入便得到最后时刻的最优 LAI_{MAX} 。

此算法实现中很重要的一步是如何得到SAIL模型对 LAI 的梯度 $\partial H(x_t)/\partial x_t$ ，这可以通过SAIL模型的切线性模式得到，而LOGISTIC模型是一个简单的非线性函数，其对待优化参数的梯度容易得到。本文设计的参数反演算法流程如图1所示。

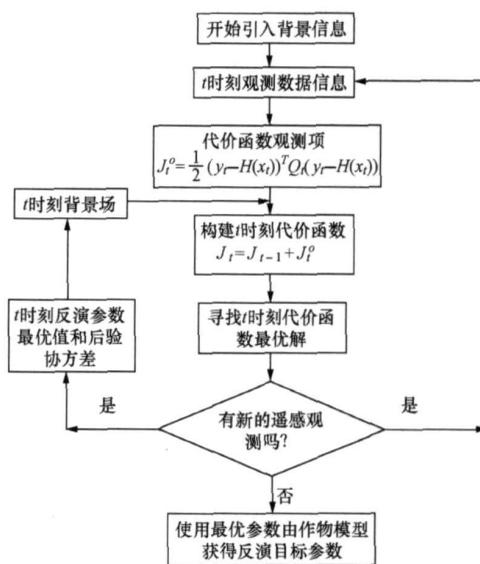


图1 地表参数同化反演流程图

2 模型处理和数据、参数误差表达

为了说明本文方法的可行性，本文采用简化的

作物生长模型和适用于水平均匀植被的辐射传输模型. 本节介绍算法实现中对模型的处理和数据、参数误差的处理方法.

2.1 模型处理

(1) 作物生长模型

LAI 是影响作物冠层光谱特征的主要冠层结构参数之一, 在作物的整个生育期内呈动态变化, 与作物生物量间存在显著相关性, 反映了作物的生长状况及其动态变化趋势.

LOGISTIC 模型是一个描述农作物群体生长的模型, 项月琴等^[10]对典型农作物 LAI 和相对积温之间的关系进行了研究, 得出冬小麦生长发育期间 LAI 和相对积温之间的关系. 林忠辉等^[11]利用修正的 LOGISTIC 方程对夏玉米的 LAI 和相对积温进行了 LOGISTIC 拟合, 得到较好的结果. 这里我们利用地物波谱数据库中 2001 年冬小麦数据采用同样的方式进行 LOGISTIC 方程拟合, 采用的方程^[11]如下:

$$RLAI = \frac{LAI}{LAI_{MAX}} = \frac{1}{1 + e^{(a_0 + a_1 * DVS + a_2 * DVS^2)}} \quad (8)$$

这里 LAI_{MAX} 是冬小麦生育期中叶面积指数最大值; DVS 为归一化积温; a_0, a_1, a_2 是未定的拟合参数; $RLAI$ 是相对叶面积指数, 是对 LAI 的归一化表达.

(2) 辐射传输模型

辐射传输模型 SAIL 表达为一个 4 流 9 参数微分方程组, 原方程^[9]如下(9)–(12)式:

$$\frac{dE_s}{dx} - KE_s = 0 \quad (9)$$

$$\frac{dE^-}{dx} + SE_s - aE^- + \sigma E^+ = 0 \quad (10)$$

$$\frac{dE^+}{dx} - S'E_s - \sigma E^- + aE^+ = 0 \quad (11)$$

$$\frac{dE_0}{dx} - wE_s - vE^- - lE^+ + kE_0 = 0 \quad (12)$$

其中 E_s, E^-, E^+, E_0 分别是冠层上界的辐照度, 冠层内部的上行辐亮度, 冠层内部的下行辐亮度和视辐亮度; $K, S, a, \sigma, S', l, v, w, k$ 是与

LAI , 观测几何等参数相关的参数系数.

在本文所采用的求解代价函数最小值的算法中要求计算反射率对 LAI 的梯度, 由于参数 LAI 在 SAIL 模型中是和其他参数变量一起构成 SAIL 模型的参数系数, 因而可以通过 SAIL 模型的切线性模式^[12]得到反射率对 LAI 的梯度. 由此对 LAI 写出模型的切线性化方程如(13)–(16)式:

$$\frac{dE_s}{dx} - KE_s - \tilde{K}E_s = 0 \quad (13)$$

$$\frac{dE^-}{dx} + SE_s + \tilde{S}E_s - aE^- - \tilde{a}E^- + \sigma E^+ + \tilde{\sigma}E^+ = 0 \quad (14)$$

$$\frac{dE^+}{dx} - S'E_s - \tilde{S}'E_s - \sigma E^- - \tilde{\sigma}E^- + aE^+ + \tilde{a}E^+ = 0 \quad (15)$$

$$\frac{dE_0}{dx} - wE_s - \tilde{w}E_s - vE^- - \tilde{v}E^- - lE^+ - \tilde{l}E^+ + kE_0 + \tilde{k}E_0 = 0 \quad (16)$$

方程中 E_s, E^-, E^+, E_0 为 SAIL 模型中由于 LAI 的扰动引起的辐亮度扰动量, $K, S, \tilde{a}, \tilde{\sigma}, \tilde{l}, \tilde{v}, \tilde{w}, \tilde{k}$ 为切线性化 SAIL 模型的参数系数, 是由 LAI 扰动而引起的 SAIL 模型相应系数扰动量. (13)–(16)式的本质含义是当 LAI 有一个微小的扰动时, 会引起相应辐亮度的微小变动, 当 LAI 的微小扰动趋近于 0 时, 辐亮度扰动与 LAI 扰动的比值极限就是辐亮度在相应 LAI 处的梯度, 而由观测的视辐亮度扰动 E_0 便可以进一步获得 SAIL 模型反射率的相应扰动量, 这个扰动与 LAI 扰动的比例极限就是反射率对 LAI 的梯度.

本文从代价函数出发根据上述切线性模式的方法获取代价函数对 LAI 的梯度. 梯度的具体求算方法可以利用自动微分技术^[13], 自动微分技术是对一个模型在计算机代码层次上直接实现模型切线性模式和伴随模式的处理技术, 它可以避免对复杂模型求切线性模式或伴随模式的复杂理论表达形式. 本文利用自动微分技术从 SAIL 模型的源代码获得 SAIL 模型的切线性代码, 这样就实现了方程(13)–(16)的代码形式.

2.2 观测数据和参数误差处理

代价函数中的观测误差项由两部分组成, 首先

是 SAIL 模型对冠层参数和冠层反射之间关系的表达存在误差, 其次观测数据本身由于仪器、天气等因素的影响同样会存在测量误差; 这两部分误差由 Q 来描述.

待优化参数先验值 u_B 本身存在的误差表达为误差协方差 B , 由于同化作用, 随着时间序列观测数据的逐次加入, B 的取值就可以由每次同化后获得的参数后验概率分布来更新. 本文采用 Tarantola^[14] 的方法获取参数后验概率密度的误差协方差. 其表达式为:

$$B_n = \left(\sum_{i=1}^n H_i^T Q_i^{-1} H_i + B_{n-1} \right)^{-1} \quad (17)$$

其中 B_{n-1} , B_n 分别为上一同化时刻和当前同化时刻的先验误差协方差. 由于本文选择单一参数 LAI_{MAX} 为同化变量, 因此这里的先验误差协方差为标量; H_i 是每个同化时刻观测算子的 Jacobi 矩阵, 本文的实验中因只有一个同化变量而为标量形式; 式中其他变量的意义同 (1) 式所示代价函数变量的说明. 这个公式表达了由待优化参数的先验信息和时间序列观测数据决定的联合概率分布对待优化参数以冠层模拟反射率为条件的条件概率分布的误差协方差.

综上所述, 本文对代价函数中的误差协方差 Q 和 B 的处理方法为: (1) 考虑观测场的误差协方差 Q 来自两个方面, 也就是由 SAIL 模型参数误差引起的模型误差, 和由观测数据 y_i 引起的测量误差; (2) 待优化参数的误差协方差随着时间进程而改变, 对其改变的实现是通过参数的后验概率分布来获得更新的 B , 也就是 (17) 式.

在计算由模型参数误差引起的 SAIL 模型误差时, 由于 SAIL 模型的输入参数都隐含在方程 (9)–(12) 的参数系数中, 直接从模型公式获取模型误差有一定的困难, 因此采用 Monte-Carlo 方法^[15] 模拟. 模拟中假定待反演目标参数的误差分布符合 Gauss 概率分布, 根据 Gauss 分布的先验信息对参数的误差分布采样 1000 次, 再带入 SAIL 模型来获得模拟反射率集合, 对这个模拟反射率集合进行统计, 由此得到模拟反射率的误差概率分布, 也就是由参数误差引起的模型误差.

3 实验数据和实验结果

本文进行参数反演研究的地表作物类型为冬小麦, 实验区域选择为 2001 年国家“九七三”项目组织的定量遥感综合实验北京小汤山冬小麦实验场的 W5 和 W6 区. 采用的数据包括冬小麦生育期间的气象数据、辐射传输模型的部分输入参数数据、冠层反射率时间序列观测数据、待优化参数的先验值及其先验方差, 均取自于“我国典型地物标准波谱数据库”^[12]. 其中冬小麦冠层反射率观测数据、观测日期、观测几何如表 1 所示.

表 1 输入数据

观测日期	太阳天			太阳方			反射率		
	顶角/(°)	位角/(°)	反射率/%	顶角/(°)	位角/(°)	反射率/%	顶角/(°)	位角/(°)	反射率/%
	(W5)	(W5)	(W5)	(W6)	(W6)	(W6)	(W6)	(W6)	(W6)
2001-04-03	35	6925	168 4442	24 8580	35 9012	166 3451	22 8557		
2001-04-13	51	3832	245 8510	35 1868	56 9234	252 3920	32 3371		
2001-04-21	28	6066	178 6781	37 3836	28 8898	189 3838	36 8212		
2001-04-30	25	6643	180 4013	35 7431	25 6724	181 6767	38 5937		
2001-05-04	32	1859	227 4792	40 0715	32 5786	228 4829	35 7586		
2001-05-09	28	6934	222 2879	39 6290	28 4344	221 4281	36 0742		
2001-05-13	25	5166	219 3802	37 2429	28 5363	134 2610	36 0400		
2001-05-23	27	3636	220 3745	34 8243	19 7723	177 7802	32 0703		

(1) 待优化参数 LAI_{MAX} 先验值和先验方差设置

冬小麦生长参数的先验值和先验方差, 均从对波谱库中的 2001 年实测数据的统计中获得, 以此构造待优化参数的先验信息, 其中: 设定参数 LAI_{MAX} 的先验值: $LAI_{MAX} B = 2.982$; 先验方差: $B = 0.364$.

(2) 观测场误差设置

观测场误差包括模型模拟误差和冠层反射率数据的测量误差, 综合估计误差设为 $Q_i = 10\% \cdot y_i$. 在模型误差中, 本文主要考虑由模型参数引起. SAIL 模型有 7 个输入参数, 其中涉及叶倾角分布的参数 2 个, 天空光比参数 1 个, 叶片反射和透射 2 个, 1 个土壤反射率参数, 这些参数可根据小汤山实验观测数据给定; 对于 LAI , 由于对 LAI_{MAX} 的估算误差会导致由 LOGISTIC 估算的 LAI 存在误差, LAI 的估算值在通过 SAIL 模型转换到反射率空间时会引起反射率的估算误差, 也就形成 Q 中的第一部分, 这实际是一个误差传递过程. 这样

SAIL 模型的误差主要由参数 LAI 的不确定性引起, 估算方法采用 2.2 中所述的 Monte-Carlo 方法. 在表 2 中列出由 LAI 的估计误差引起的反射率误差

百分比, 由于 LOGISTIC 模型建立了相对积温与 LAI 之间的关系, 故在表 2 中也列出各个时刻的相对积温值.

表 2 SAIL 模型误差评估

观测时间步	1	2	3	4	5	6	7	8
归一化积温值	0.6074	0.6905	0.7817	0.8616	0.9166	0.9854	1.0722	1.3290
估算的 LAI 值	0.3002	1.4278	2.5326	2.8253	2.8912	2.9232	2.9308	2.5266
LAI 的估算误差	0.0037	0.0834	0.2626	0.3267	0.3422	0.3498	0.3516	0.2613
模拟反射率值	0.1846	0.3201	0.3823	0.3920	0.3940	0.3949	0.3951	0.3821
由 LAI 引起的反射率标准差	0.0100	0.0212	0.0158	0.0141	0.0136	0.0134	0.0132	0.0158
由 LAI 引起的反射率相对误差/%	5.42	6.62	4.13	3.6	3.45	3.39	3.34	4.14

由表 2 可见, 由 LAI 不确定性引起的模拟反射率误差基本在反射率观测值的 3%—6% 之间. 因而在本文中我们将由 LAI 引起的 SAIL 模型误差设为观测值的 5%, 这样做的一个前提是认为 SAIL 模型本身是没有理论上的不确定性, 这是因为本研究使用的反射率数据没有热点方向的观测, 而且 Verhoef 已经验证了 SAIL 模型模拟反射率与实际测量值的一致性^[9], 因而可以考虑其模拟的反射率误差仅仅来源于参数的不确定性.

Q_c 中的另外一部分则是冠层反射率数据的测量误差, 主要来源于仪器本身的精度、操作人员的不规范测量以及环境因素的影响等, 这样的误差在本文中并没有进行定量的评估, 其原因是测量人员的不规范操作和环境因素的影响等很难确定, 而且测量误差对算法本身的影响是在相对的程度, 同样由于本文的一个重点是放在所设计的数据同化算法验证上面, 我们假设冠层反射率的测量误差与 SAIL 模型的模拟误差具有对称性, 即也假设测量反射率的相对测量误差是 5%. 经过这样的处理, 最终我们设定 Q_c 为观测反射率数据的 10%. 采用上述同化算法, 用冠层反射率时间序列同化 LAI_{MAX} 的结果列于表 3 中.

表 3 引入波谱数据库统计值作为 LAI_{MAX} 先验信息的

LAI_{MAX} 同化结果

同化步数	1	2	3	4	5	6	7	8	真实 LAI_{MAX}
W5 区	3.002	3.046	3.053	3.060	3.070	3.079	3.088	3.097	3.153
W6 区	3.019	3.011	2.936	2.870	2.817	2.761	2.724	2.692	2.684

由表 3 可见, LAI_{MAX} 在各个同化步处的最优值随着时间序列观测数据的加入而逐渐接近真实值, 同化步数 1—8 代表引入的观测反射率个数. 在同化的初始时刻, 也就是仅有一个观测反射率数据时, 利用 LAI_{MAX} 的先验信息和反射率观测数据提供的信息, 同化得到的最优 LAI_{MAX} 对于 W6 区是 3.019, 离真实值较远. 而在时间序列上引入 8 个观测反射率数据后, 同化得到的最优 LAI_{MAX} 值 2.694 已经十分接近真实值 2.684. 这一结果说明本文算法的两个优点: (1) 时间序列观测反射率数据的引入提供了更多的信息, 有效地改进了参数优化结果; (2) 用前一时间段的同化结果更新参数反演先验信息的不确定性, 使各个同化步处的先验误差不确定性逐渐降低.

将由 8 个同化步得到的 LAI_{MAX} 值引入作物生长模型 LOGISTIC, 得到 8 条 LAI 随年积日变化的反演趋势线, 连同各个年积日处测量的真实 LAI 数据一同绘制于图 2, 3, 分别对应 W5 和 W6 实验区, 图中横坐标为年积日. 在图 2, 3 中同时也绘出基于单一时刻观测反射率分别反演的 LAI 结果, 如图中标注为 SINGLE 的散点, 以便与本文引入作物生长模型和时间序列观测反射率数据 LAI 反演结果做比较.

图 2, 3 中标注为 TRUE 的粗线分别代表 W5 和 W6 区的 LAI 地面测量值, 其余标注为 1—8 的折线依次代表将 1—8 次观测反射率加入同化进程得到的 LAI 反演结果. 由图 2 可见 8 条反演 LAI 曲线几乎重合, 而图 3 中的 8 条反演 LAI 曲线有一

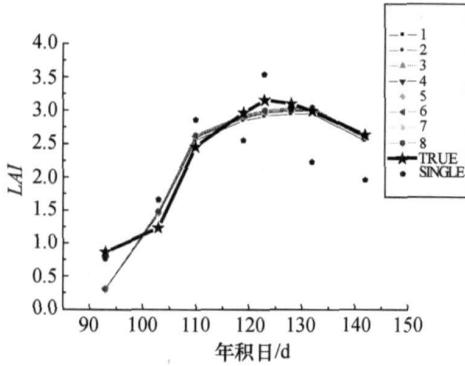


图 2 小汤山 W5 区的 LAI 反演结果

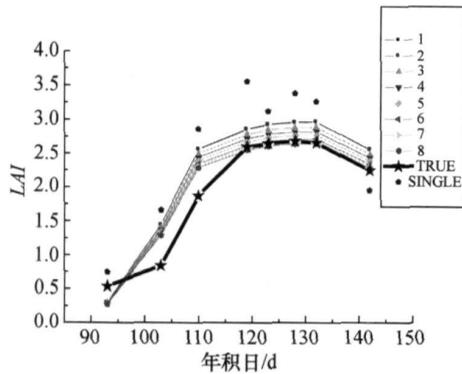


图 3 小汤山 W6 区的 LAI 反演结果

表 4 没有引入波谱数据库 LAI_{MAX} 先验信息时 W6 区 LAI_{MAX} 的同化结果

同化步数	1	2	3	4	5	6	7	8	真实 LAI_{MAX}
$LAI_{MAX}B=3.8$	3.887	3.641	3.363	3.148	2.991	2.887	2.797	2.731	2.684
$LAI_{MAX}B=1.8$	1.87	1.93	1.984	2.104	2.204	2.32	2.376	2.439	2.684

从表 4 可以看出, 在同化的初始时刻因为缺乏良好的先验信息, 导致 LAI_{MAX} 的同化结果离真实值很远, 明显差于表 3 所示采用良先验信息的结果. 但是随着同化进程到引入 8 个观测反射率时, LAI_{MAX} 值已经接近真实值, 可见时间序列信息的作用非常明显, 由此可见, 在同化过程中引入时间序列信息的必要性.

将 8 个时刻同化得到的最优 LAI_{MAX} 引入作物生长模型 LOGISTIC, 由此得到 8 条反演 LAI 变化趋势曲线, 结果如图 4, 5. 从图 4, 5 可见, 当先验信息不足时, 无论是高估还是低估背景先验信息, 时间序列观测反射率的引入都能够带来足够的有用信息, 这种信息的补充能够明显减小由于先验信息不足而引起的反演误差, 提高参数反演精度.

定分离性, 但相同的趋势是: 随着时间序列观测反射率的不断加入, LAI 反演结果曲线逐渐趋近测量值. 由于 LOGISTIC 模型的拟合误差, 可以看到 LAI 反演结果在初始一段时间内(年积日 95—105) 有较大的偏差, 但在曲线的后半段, 其总体效果比较好, 由此可见过程模型本身误差所带来的影响. 对比基于单一时刻观测反射率的反演结果, 可以发现散点跳跃性较大, 在图 2, 3 中都存在 LAI 值的单谷, 分别对应于图 2 中 119 d 处, 图 3 中的 123 d 处, 其与作物生长过程中 LAI 连续变化的规律不符. 而通过引入作物生长模型 LOGISTIC 的同化反演则实现了用 LAI 自身变化规律对不合理的反演 LAI 值进行合理的校正.

为了突出说明引入时间序列观测数据和作物生长模型对同化反演结果的作用, 对 W6 区重新进行 2 组同化计算, 唯一的区别是不再采用波谱库中 LAI_{MAX} 的先验信息, 因而这时没有先验信息可以利用, 分别假设对参数先验信息值高估和低估, 这两种情况都是对先验信息认识不够的情况, 试验数据为: (1) 设定高估先验信息值 $LAI_{MAX}B=3.8$, $B=0.9$; (2) 设定低估先验信息值 $LAI_{MAX}B=1.8$, $B=0.9$. LAI_{MAX} 的同化结果如表 4:

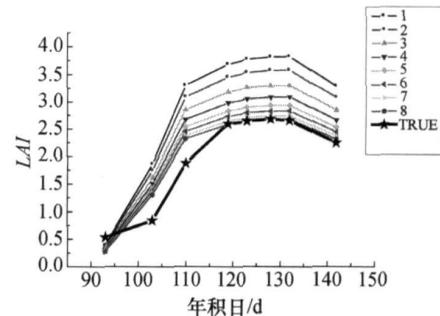


图 4 高估背景时 LAI 的反演结果

图 6, 7 分别表示具有高估先验信息和良先验信息情况下同化进程中待优化参数 LAI_{MAX} 的演进过程. 图中曲线代表同化过程中每个同化步处的

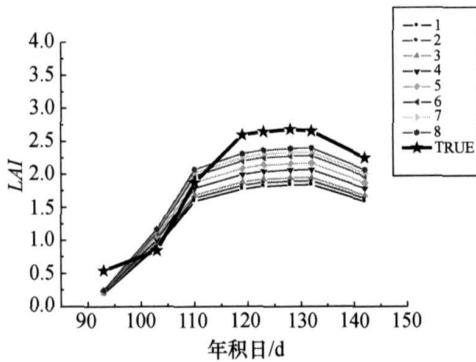


图 5 低估背景场时 LAI 的反演结果

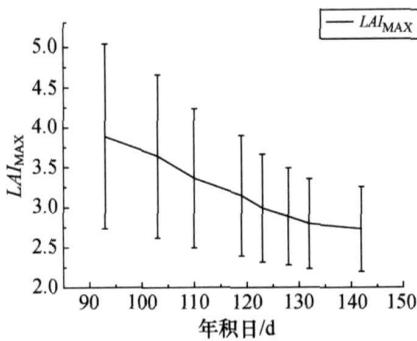


图 6 高估背景时 LAI_{MAX} 的方差演进

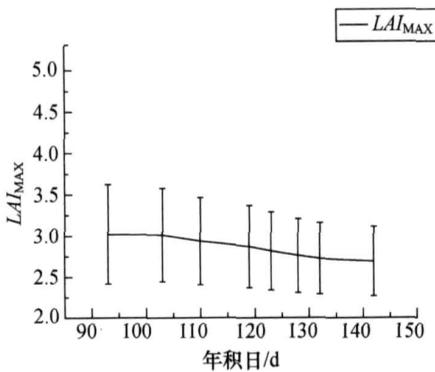


图 7 良背景场时 LAI_{MAX} 的方差演进

LAI_{MAX} 最优值, 同时曲线上在每个同化步处标出 LAI_{MAX} 的标准差区间. 从图 7 可见, 在具有良先验信息时, 时间序列观测反射率数据虽然可以降低 LAI_{MAX} 的不确定性, 但并不如图 6 中明显, 究其原因是因为先验信息已经具有足够精度, 时间序列观测反射率仅仅能够对 LAI_{MAX} 进行一定程度的优化. 而在图 6 中, 引入多时相时间序列观测反射率的作用很明显, 此时虽然背景先验信息不足, 但通过引

入时间序列观测反射率数据, 使得 LAI_{MAX} 逐步得到显著修正, 表现为演进过程中 LAI_{MAX} 不确定性空间减小的趋势. 由此, 从实验结果可以看到, 采用时间序列的数据同化策略对降低待优化参数不确定性的作用.

4 结论与讨论

本文利用地物波谱库中参数先验知识所提供的信息, 利用作物生长过程模型, 结合时间序列的冠层反射率观测数据, 采用变分数据同化方法对作物生长参数 LAI 进行反演. 借鉴传统基于先验信息的遥感反演方法实现了多时相观测数据和作物生长模型的信息融合. 研究证明采用先验知识构造一个良好的先验信息对于由观测数据获得更好 LAI 反演结果的重要性; 也验证了采用时间序列观测数据可以削弱先验信息不足带来的不良影响, 同时可以提高参数反演精度; 同时也说明了在遥感模型反演中引入过程模型的约束作用.

从遥感观测提取地表植被参数对农业生态监测, 环境评估, 以及陆地表面过程模拟等领域具有重要的应用价值. 在将本文方法应用于实际的过程中则可以充分利用高时间分辨率的遥感观测数据, 或不同传感器对同一地区的重复观测. 而基于先验信息的遥感反演方法在实际遥感反演过程中的很大一部分难点是先验信息的获取, 而我国典型地物波谱数据库的建立则为传统基于先验信息的遥感反演提供了可利用的数据源, 也为本文方法在实际中的应用提供了有效的先验信息源.

本文数据同化方法反演地表参数中有一个重要的问题就是代价函数中观测数据误差的设定, 本文就此进行研究, 给出了一种确定误差设定值的方法, 并得出了方法验证结果. 在此进一步讨论当测量误差的假设不成立时对反演结果的影响, 这里涉及 3 个问题: (1) Q_t 的测量误差部分大于或小于 5% 的设定值时对反演结果的影响; (2) 当时间序列观测中的一个或几个观测存在误差较大的观测, 而 5% 的测量误差没有足够描述这个误差时对反演结果的影响; (3) 当一个或几个观测存在较大的观测误差, 而在数据质量控制过程中注意到这个误差时对反演结果的影响. 对应于上述 3 种情况我们增加 3 组实验, 而对后两种情况则主要观察时间序列观

测其中的一个观测数据错误评估误差的情况, 这里采用对第 5 次观测反射率加人为扰动, 对应数据为表 1 中的 2001 年 5 月 4 日的反射率数据. 在这 3 组实验中, 对待优化参数 LAI_{MAX} 先验信息采用了高估先验信息时的数据, 这是因为由表 4 和图 4 的结果可见高估先验信息时同化结果的分异性更明显. 3 组实验的 LAI_{MAX} 同化结果列于表 5 中. 表 5 中第 2 行是设定观测数据的测量相对误差为 10% 时的反演结果, 第 3 行是第 5 次观测人为增加 10% 的扰动, 但没有进行正确误差评估时的反演结果. 第 4 行的同化结果则是第 5 次观测加上 10% 的人为扰动, 但正确评估误差时的同化结果. 为便于比较, 高估先验信息时的 LAI_{MAX} 原始同化结果由表 4 重新抄录到表 5 的最后一行.

由表 5 可见, 当假定测量误差达到 10% 时, 时间序列观测能够改进 LAI_{MAX} 的同化反演结果, 但在各个同化步处的同化结果相对于原始同化结果已经出现偏移, 但在第 8 个同化步处的 LAI_{MAX} 同化结果还是能够反映真实值的, 但已经不如设定 5% 测量相对误差时的同化结果好, 可见采用 5% 的测量相对误差是合适的. 针对于第 2 种情况, 第 5 次观测被人为加上 10% 的扰动, 这个扰动的施加导致第 5 次观测是一个错误的观测. 从表 5 可见, 由于在同化的过程中并没有进行合适的数据质量控制, 在前 4 步的同化过程中, 同化结果还可以按照正确的路线行进, 但在第 5 个同化步处, 同化结果没有

什么实质的变化, 而能够维持 3.145 的原因则是前 4 步同化信息积累的结果. 由于这个错误观测的引入, 导致最后第 8 步的同化结果出现较大偏差, 但是 8 个同化步骤中观测数据信息积累对改进反演的有效作用同样可以见到. 从这个实验可以发现对观测数据进行质量控制的必要性. 针对于上述第 3 种情况, 如果在同化过程中注意到第 5 次观测有较大的误差, 而且这个误差得到正确评估, 则第 5 次观测在整个代价函数中所占的比例将减少, 第 5 次观测对 LAI_{MAX} 的贡献也将减少, 从表 5 中的结果可见其影响, 这种影响实际上是来源于第 5 次观测提供信息的减少.

从以上的 3 组实验和讨论可以发现, 本文所采用的数据同化方法对时间序列观测可以实现信息积累, 这是本文的研究重点; 同时, 对观测数据进行质量控制也是必要的, 错误的观测数据会引起较大的反演误差.

LOGISTIC 模型是一个简化的作物生长模型, 采用这个模型的原因是为了简化算法, 以便验证算法的可行性. 在进一步的工作中我们已经采用机理化的作物生长模型 CERES 实现了北京地区冬小麦的 LAI 反演, 实现了本文算法的实际应用. 研究表明, 基于不同的作物生长模型所选取的待优化参数 u 也会有相应的变化, 例如针对 CERES 模型, 待优化的参数可以选取平均叶热间距, 施肥量, 灌溉量等参数.

表 5 不同观测误差评定现象对反演结果的影响

同化步数	1	2	3	4	5	6	7	8	真实 LAI_{MAX}
测量误差采用 10% 的反演结果	3.786	3.692	3.505	3.334	3.175	3.053	2.950	2.869	2.684
第 5 次观测误差较大, 没有正确评估; $LAI_{MAX}B=3.8$	3.887	3.641	3.363	3.148	3.145	3.148	3.140	3.101	2.684
第 5 次观测误差较大, 正确评估; $LAI_{MAX}B=3.8$	3.887	3.641	3.363	3.148	3.095	2.972	2.853	2.789	2.684
高背景时原始同化结果; $LAI_{MAX}B=3.8$	3.887	3.641	3.363	3.148	2.991	2.887	2.797	2.731	2.684

参 考 文 献

- Li XW, Wang JD, Hu BX, et al. On utilization of prior knowledge of remote sensing model. Science in China, series D, 1998, 41(6): 580-586
- 王锦地, 李小文. 地物波谱数据库与地表参数的遥感反演. 遥

感学报, 2004, 8(增刊): 4-7

- Le Demit FX, Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. Tellus (series A), 1986, 38: 97-110
- Clevers JGPW, Vonder OW, Raymond EE, et al. Using SPOT data for calibrating a wheat growth model under mediterranean

- conditions. *Agronomie*, 2002, 22: 687—694
- 5 Launay M, Guerif M. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture, Ecosystems and Environment*, 2005, 111: 321—339
 - 6 Doraiswamy PC, Sinclair TR, Hollinger S, et al. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sens Environ*, 2005, 97: 192—202
 - 7 Jarlan L, Mangiarotti E, Mougou E, et al. Assimilation of SPO T/ VEGETATION NDVI data into a sahelian vegetation dynamics model. *Remote Sens Environ*, doi: 10.01016/j. rse. 2007. 02. 041
 - 8 Arken G. *Mathematical Methods For Physicists*. Orlando: Academic Press, 1985, 945—950
 - 9 Verhoef W. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens Environ*, 1984, 16: 125—141
 - 10 项月琴, 王锦地, 李小文, 等. 二向反射遥感中冬小麦植被组分和土壤特性的季相变化. *遥感学报*, 2000, 4(增刊): 90—100
 - 11 林忠辉, 项月琴, 莫兴国, 等. 夏玉米叶面积指数增长模型的研究. *中国生态农业学报*, 2003, 11(4): 69—72
 - 12 黄思训, 伍荣生. *大气科学中的数学物理问题*. 北京: 气象出版社, 2001, 466—472
 - 13 程 强, 张林波, 王 斌. 模式伴随化的基本规则及其代价分析. *中国科学, E 辑*, 2004, 34(6): 601—618
 - 14 Tarantola, A. *Inverse Problem Theory: Methods For Data Fitting and Model Parameter Estimation*. New York: Elsevier Science Publishing Company Inc, 1987, 68—72
 - 15 Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geo-Physical Research*, 1994, 99(5): 10, 143—10, 162

(上接第 873 页)

四、说明:

(一) 对于有很强的科研学术工作背景, 业绩突出, 又有丰富的期刊编辑、管理工作经验的, 可酌情考虑放宽年龄条件。

(二) 应聘人员被正式聘用后, 薪酬待遇从优。

(三) 要了解国家自然科学基金委员会和科学基金杂志社有关情况, 请查询 www.nsf.gov.cn, 欢迎有意者报名应聘。报名时请填写《国家自然科学基金委员会科学基金杂志社工作人员申请表》, 通过电子邮件或书面材料寄下列地址:

通讯地址: 北京市海淀区双清路 83 号 国家自然科学基金委人事局干部处

邮政编码: 100085

截止日期: 2008 年 8 月 30 日。

咨询电话: 010-62326928, 010-62326934

电子邮件: tanglh@nsf.gov.cn wangwz@nsf.gov.cn

报名截止后, 招聘工作小组研究决定答辩人选, 将以电话或电子邮件方式通知答辩。

附件: 国家自然科学基金委员会科学基金杂志社工作人员申请表 <http://www.nsf.gov.cn/nsfc/cen/zhao pin/fj/fj2008070201.doc>

国家自然科学基金委员会

人事局

2008 年 7 月 1 日