

文章编号: 1007-4619 (2000) 02-0090-05

# 陆地表面粗糙度和土壤湿度多维参数 同时反演的遗传算法

王悦泉, 金亚秋

(复旦大学波散射和遥感中心, 上海 200433)

**摘 要** 用遗传算法, 从多角度后向散射观测值同时反演陆地表面粗糙度(小尺度起伏方差与相关长度、大尺度起伏坡度)和土壤体湿度的多维特征性参数。粗糙陆地表面散射用 Kirchhoff 近似的稳相法与微扰法相结合的双尺度模型计算, 以构造代价函数。遗传算法反演结果与车载多角度遥感时的陆地实测值作了比较, 取得了十分良好的结果。为遗传算法在遥感反演中的新应用提供了实例。

**关键词:** 遗传算法; 陆地粗糙度与湿度; 反演

**中图分类号:** TP721.2/S15 **文献标识码:** A

## 1 引 言

确定土壤湿度是对地遥感的最主要目标之一。土壤含水量(单位体积中含水量定义为体湿度)的变化改变土壤的介电常数, 从而影响遥感观测的后向散射系数( $\sigma$ )和热辐射亮度温度( $T_B$ )。由于自然陆地表面一般总是起伏不平的, 这一表面的粗糙度同样也影响遥感观测的 $\sigma$ 和 $T_B$ 。因此, 土壤的湿度与地表的粗糙度是相互关联的参数组合, 它们共同影响着观测值 $\sigma$ 和 $T_B$ 。如何从遥感观测中更好地分别反演出土壤湿度和地表粗糙度参数, 一直是遥感研究和应用的一个重要任务。

我们曾提出一种网格法<sup>[1]</sup>和人工神经网络方法(ANN)<sup>[2]</sup>来反演地表的粗糙度和土壤湿度, 取得了很好的效果。当然, 网格法要确定土壤的构造类型, 以构造不同的网格, 而 ANN 方法则需要较长的训练时间和大量的训练数据。

近些年来, 模仿自然界生物优胜劣汰优选法则的遗传算法, 在机器学习、过程控制、优化设计等工程领域得到了广泛的应用<sup>[3,4]</sup>。生物的遗传密码是基因, 基因组成染色体, 一种染色体对应着一个生物个体, 不同染色体的组合则代表着一个生物群体。适应自然条件的生物体才能得以生存发展和繁殖,

这样经过多次优胜劣汰后产生的将是优化的后代。生物界这一法则应用到工程领域就形成了遗传算法, 该算法将参数编码形成为基因, 基因组合形成染色体, 对基因及其组合反复进行自然选择、克隆、变异、杂交等步骤, 最终便可达到最佳的参数组合。

本文提出用这样的遗传算法, 通过遥感观测的后向散射系数, 来同时反演土壤的湿度和地表的粗糙度多维参数的组合, 要求所得的多维参数组合是最能符合地表实况的最佳结果。

## 2 多维参数反演的遗传算法

自然界包含形形色色的生物, 决定生物各自特性的主要遗传密码是基因。John Holland 在 1992 年陈述了他根据生物遗传的优化法则所提出的人工遗传算法<sup>[5]</sup>, 该算法通过模拟生物遗传优化, 来求得特定问题中的最佳选择。近些年来, 随着计算机技术的迅速发展和计算速度的提高, 遗传算法得到了新的重视, 并在机器学习、过程控制、工程设计优化等多个领域内得到了十分广泛和重要的应用, 成功地解决了许多优化设计问题。它象人工神经网络(ANN)方法一样, 展示了自然界生物体进化选择过程对人类进行工程优化设计的思维启迪。遗传算法的简单流程见图 1。

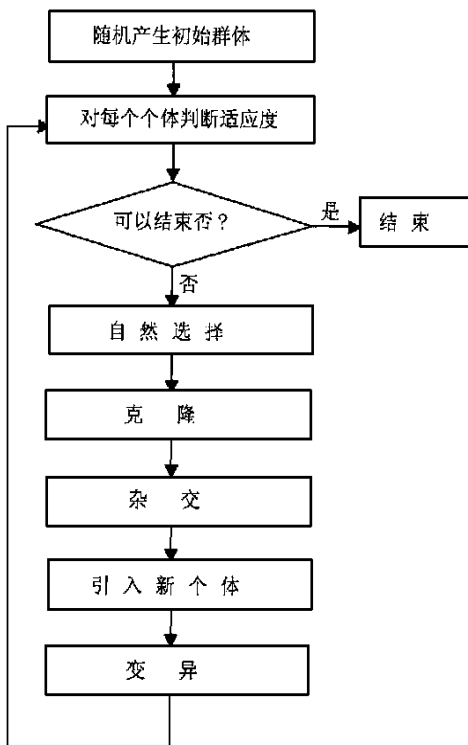


图1 遗传算法流程图

Fig. 1 Flow chart of a genetic algorithm

基因的不同组合形成不同的染色体,一种染色体对应着一个生物个体,而生物群体则可由特定数目的染色体表示。典型的遗传算法须将每个参数编码形成基因,近几年,在求解高维或复杂优化问题时大多使用实数编码,这是由于实数编码表示比较自然,而且易引入相关领域知识。

遗传算法首先随机产生初始的生物群体,至于群体规模的大小,要根据待解决的问题本身来确定,个体数目越多,则搜索范围越广,易获得全局最优,但迭代时间长;个体数目少则易导致算法不能或很慢才能收敛至最优。

然后对生物群体实施自然选择算子。在自然选择算子的过程中,有多种不同的选择策略,比较常见的有轮盘式选择策略<sup>[6]</sup>和确定式选择策略<sup>[7]</sup>。我们在程序中采用的是确定式选择策略,即根据各条染色体所对应的适应度的大小,将染色体排序,只保留适应度大的一部分染色体,使之成为下一代的父体,其余的则被淘汰。对存活下来的父体,可以对其进行克隆操作,以增加优秀个体在群体中的比例,提高群体的平均适应度。对存活下来的父体,使其两两杂交,产生具有父代个体优良特性的新个体,并且维持群体规模。实数编码基因的杂交方式常见的有离散杂交、算术杂交等。我们在算法的实现中,采用的

是算术杂交,即对父体用权重进行算术相乘相加的方法产生新的个体。在群体的每一代中,还须通过适量的变异或引入新个体的方法来保证生物群体的种类多样性,从而减小陷入局部最优的可能性。变异及引入新个体的比例不能太大,太大则会破坏已有的良好的染色体,导致算法向随机搜索法方向发展。好的变异算子往往能大大加快程序的收敛速度。变异算子的设计要根据参数本身的特征来确定。

这样,每一代群体通过衡量其适应度来实现优胜劣汰,通过克隆、交配、变异、引入新个体等遗传操作来产生新的群体,如此反复,最终达到最优群体。

通过以上对遗传算法的分析,可以发现,与一般建立在求梯度基础上的传统的搜索方法相比,由于该算法是以群体为单位展开搜索,并且由于变异和引入新个体等遗传算子的存在,所以其具有一个显著的优点,即不易陷入局部最优;而且,它不要求微分等复杂的数学计算,能很好地解决离散型参数和参数数目很多时的优化问题。

本文中我们提出用遗传算法来同时反演土壤湿度和地表粗糙度。将所需反演的体湿度  $m_v$  和粗糙度(小尺度的起伏方差  $\sigma_1$  及相关长度  $l_1$ 、大尺度的平均坡度  $s = \sqrt{2} \sigma / l$ , 这里,  $\sigma, l$  分别为大尺度的起伏方差和相关长度)作为基因。它们的一种组合作为一种染色体。几种不同染色体的组合,就代表着一个生物群体。

### 3 粗糙地面的双尺度散射模型

土壤的体湿度( $m_v$ )定义为土壤单位体积的含水量,含水量的多少可以改变土壤的介电常数  $\epsilon (= \epsilon' + i\epsilon'')$ 。影响土壤介电常数  $\epsilon$  的还有土壤结构,如粘土和沙土的含量,水份从束缚状态到自由状态的过渡湿度的确定等等。微波频段的土壤介电常数可由经验性公式计算<sup>[8]</sup>。

地表面的随机起伏可由起伏方差  $\sigma$  和随机面在水平方位上的统计相关长度  $l$  来定义。当粗糙地表满足大尺度起伏条件

$$kl > 6 \text{ 与 } l^2 > 2.76 \sigma \lambda \quad (1)$$

时( $k$  为波数,  $\lambda$  为波长),可采用 Kirchhoff 切平面近似(KA)的稳相法<sup>[8]</sup>, 得到同极化( $pp = hh$  或  $vv$ )后向散射系数为:

$$\sigma_{pp}^{KA}(\theta) = \frac{|R_p(0)|^2}{2s^2 \cos^4 \theta} \exp \left\{ -\frac{\tan^2 \theta}{2s^2} \right\} \quad (p = h, v) \quad (2)$$

其中  $\theta$  为入射角,  $s$  为大尺度起伏表面的平均坡度, 对 Gauss 表面有  $s = \sqrt{2} \sigma_1 / l_1$ ;  $R_h(0)$ 、 $R_v(0)$  为法向 ( $\theta = 0$ ) 入射平滑表面的水平极化与垂直极化的 Fresnel 反射系数。

当粗糙地表为微小起伏, 满足

$$k\sigma_1 < 0.3 \text{ 与 } \sqrt{2} \sigma_1 / l_1 < 0.3 \quad (3)$$

时(其中  $\sigma_1$ 、 $l_1$  为小尺度粗糙面的均方根高度和相关长度), 由微扰法(SPA)得到同极化( $pp = hh, vv$ )后向散射系数为<sup>[8]</sup>

$$\sigma_{pp}^{\text{SPA}}(\theta) = 8k^4 \sigma_1^2 \cos^4 \theta |a_p|^2 W(2k \sin \theta, 0) \quad (4)$$

$$\text{其中 } a_h = \frac{\cos \theta - \sqrt{\epsilon_r - \sin^2 \theta}}{\cos \theta + \sqrt{\epsilon_r - \sin^2 \theta}}$$

$$a_v = (\epsilon_r - 1) \frac{\sin^2 \theta - \epsilon_r(1 + \sin^2 \theta)}{[\epsilon_r \cos \theta + (\epsilon_r - \sin^2 \theta)^{1/2}]^2} \quad (5)$$

这里的  $W(k_{\perp} = 2k \sin \theta, k_z = 0)$  是表面起伏相关函数的傅立叶变换, 即小尺度起伏波谱。

而更一般的, 对于有大小不同起伏尺度的粗糙表面, 可以采用双尺度近似, 即将反射率有修正的大尺度起伏 KA 解, 和在大尺度坡度上作平均的小尺度起伏 SPA 解的独立迭加<sup>[8]</sup>

$$\sigma_{pq}(\theta) = \sigma_{pq}^{\text{KA}}(\theta) + \langle \sigma_{pq}^{\text{SPA}}(\theta) \rangle \quad p, q = v, h \quad (6)$$

而对大尺度起伏作了平均的 SPA 解写为

$$\langle \sigma_{pq}^{\text{SPA}}(\theta) \rangle = \frac{1}{2\pi} \int_{\pi/2}^{-\pi/2} \{ \sigma_{pq}^{\text{SPA}}(\theta') |_{\theta_n = \tan^{-1} s} + \sigma_{pq}^{\text{SPA}}(\theta') |_{\theta_n = -\tan^{-1} s} \} d\varphi_n \quad (7)$$

其中  $\theta' = \cos^{-1}[\cos \theta_n \cos \theta + \sin \theta_n \cos \varphi_n \sin \theta]$ 。

图 2 给出在双尺度模型近似下, 不同湿度和不同粗糙度下的  $\sigma_{hh}$ 。仿照已有的实验<sup>[9]</sup>, 土壤含沙量为 0.47, 含粘土量为 0.15, 测量时温度为 27°C。粗实线表示体湿度  $m_v = 0.1$ ,  $k\sigma_1 = 0.2$ ,  $kl_1 = 2.0$ ,  $s = 0.2$  时的后向散射; 细实线表示粗糙度不变, 而含水量增大为  $m_v = 0.4$  时的后向散射; 划线表示的是含水量不变, 而粗糙度减小为  $k\sigma_1 = 0.05$ ,  $kl_1 = 2.0$ ,  $s = 0.2$  时的后向散射。从图中可以看出体湿度和表面粗糙度对多角度后向散射的特征性影响。

## 4 遗传算法反演参数

我们用双尺度近似作粗糙表面散射的正向模拟计算, 用 30°–70° 五个角度的水平极化后向散射系数的实测数据  $\sigma_{hh}^T(\theta)$  (上标  $T$  表示实测), 通过遗传算法来反演参数  $k\sigma_1$ ,  $kl_1$ ,  $s$  和土壤的介电常数  $\epsilon'$

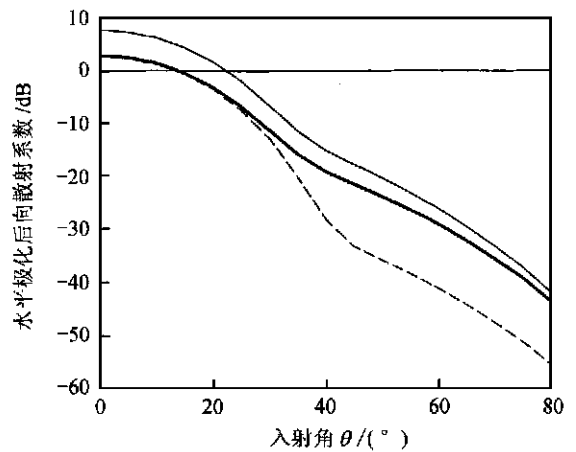


图 2 不同湿度和粗糙度的水平极化后向散射

Fig. 2 hh-polarized backscattering from land surface with different roughness and wetness

$\epsilon''$ , 再由土壤介电常数  $\epsilon$  和  $\epsilon - m_v$  经验公式反推体湿度  $m_v$ 。然后, 用反演出来的 ( $\sigma_1$ ,  $l_1$ ,  $s$ ,  $m_v$ ) 进行双尺度散射模拟, 计算  $\sigma_{hh}^R(\theta)$  (上标  $R$  表示反演后的计算值)。构造代价函数为

$$c = \sum_{\theta} | \sigma_{hh}^T(\theta) - \sigma_{hh}^R(\theta) |^2 \quad (\text{dB}) \quad (8)$$

由代价函数值  $C$  的大小来确定反演参数组合的适应度, 代价函数值  $C$  越小, 适应度越大。

在遗传算法反演土壤多维参数时, 将待反演的  $M$  个参数(本文中  $M=5$ )  $k\sigma_1$ ,  $kl_1$ ,  $s$  和 ( $\epsilon'$ ,  $\epsilon''$ ) 作为基因, 采用实数编码策略, 每种  $k\sigma_1$ ,  $kl_1$ ,  $s$  和 ( $\epsilon'$ ,  $\epsilon''$ ) 的组合作为一个染色体。

首先随机产生群体规模为  $N$  的染色体组(即  $N$  种  $k\sigma_1$ ,  $kl_1$ ,  $s$  和 ( $\epsilon'$ ,  $\epsilon''$ ) 的参数基因组合, 我们选  $N=8$ )。每个参数的产生范围根据其物理意义来确定, 我们限定  $k\sigma_1$ ,  $kl_1$ ,  $s$  和 ( $\epsilon'$ ,  $\epsilon''$ ) 的定义区间分别为  $k\sigma_1 \in [0, 1]$ ,  $kl_1 \in [0, 10]$ ,  $s \in [0, 1]$ ,  $\epsilon' \in [3, 30]$ ,  $\epsilon'' \in [0, 10]$ 。然后根据公式(9)分别计算群体中每个染色体代价函数, 再按适应度大小对染色体进行排序, 舍弃适应度较小的  $N/2$  个染色体。

对存活下来的  $N/2$  个染色体, 将其中适应度最大的两个染色体进行算术杂交, 即随机产生一对取值范围为(0, 1)的小数, 作为权重因子分别与两条最佳染色体相乘后相加, 产生一个新的染色体。如此执行  $N/2$  次, 则产生  $N/2$  个具有父代优良特性的后代个体, 群体规模保持不变。

对该新的群体再一次计算其中各条染色体的代价函数, 然后进行自然选择, 舍弃掉适应度最差的  $m$  个染色体(在我们的计算程序中, 令  $m=2$ ), 并随机产生  $m$  个新的染色体加入群体中, 使群体规模依然维持在  $N$ 。这里更新的比率  $m/N$  不宜过大, 以免算法向随机搜索法靠拢。

变异就是随机对某些基因进行突变，在实数编码时，即对某些参数基因的值进行一定范围的增加或减小。变异算子的设计必须根据待变异参量自身的特点。变异幅度太大，会跨过最优区；变异幅度太小，会导致收敛很慢。不仅如此，变异的幅度还必须参考群体的演化进度，即当群体的平均适应度较小时，变异幅度应大些；当群体的平均适应度较大、接近最优时，变异幅度应随之变小。例如，我们对  $kl_1$  的变异幅度设计为

$$|kl_1(i+1) - kl_1(i)| = 10 \times \frac{1}{20 + \frac{1}{R^{C_0/t(i)}}} \quad (9)$$

上式中  $i$  表示群体的迭代次数，10 为  $kl_1$  的定义区间大小，20 是为了限制变异最大幅度， $C_0$  为程序终止时的代价函数值大小， $t(i)$  为第  $i$  代群体中该  $kl_1$  所在的染色体所对应的代价函数值， $R$  为  $[0, 1]$  内的随机数。以上这个变异算子较好地满足了对上述对变异幅度的要求。实验证明，好的变异算子对于算法的快速收敛作用颇大。

重复以上步骤，直至群体中的最佳染色体的代价函数小于 3dB，且反演所得参数代入模型推得的后向散射系数在小角度时不小于测量值时结束。

对实验数据点的选取不能过少，要能在大致的范围内定出散射系数与入射角的关系；同时通过合理的判断，剔除冗余点和不合理点，一则减少不必要的运算，二则以免为了迎合这些冗余点，反而在合理点上造成误差。选取实验点，还必须联系物理模型本身的特点。由于在小角度时观测易受环境的影响，所以我们采用  $30^\circ - 70^\circ$  的五个实测值来反演；而小角度时的实测值仅作为参考点，要求由反演出的参数计算得到的在小角度时的  $\sigma_{hh}^r(\theta)$  不小于真实的测量值  $\sigma_{hh}^t(\theta)$ 。事实上，也不可能存在一个最佳角度来反演多维参数。

反演过程中，还须考虑一些算法之外的物理条件的限制，以避免产生不符合客观实际的反演结果。比如，要遵守模型的适用条件(见公式(1)、(3))，而利用介电常数的实部  $\epsilon'$  和虚部  $\epsilon''$  之间的联系，可避免无谓的搜索、加快收敛速度、使反演结果更加合理。

我们用 1986 年 6 月 11 日在瑞士 Marfeldingen<sup>[9]</sup> (BA2) 用 3.1 GHz 和 1986 年 8 月 14 日与 15 日在瑞士 Niederried<sup>[9]</sup> (BA3) 用 4.6 GHz 所测得的  $\sigma_{hh}^t(\theta)$  来进行反演。每次反演在 586 PC 机上约需几分钟。实验测量结果中给出的已知量有当地当时的小尺度的  $kl_1$ 、大尺度的  $k\sigma$  及土壤体湿度  $m_v$ ，我们将反演结果与当地当时实测值作了比较，其中的符合程度是令人满意的。所得结果见图 3 至图 5。图中标有方块的线段表示 BA2，标有三角的线段表示 BA3；图

中的直线段表示反演结果的方差，中间的黑点为反演的平均值。

图 3、图 4 是所得的粗糙度  $kl_1$ 、 $k\sigma$  的反演结果。

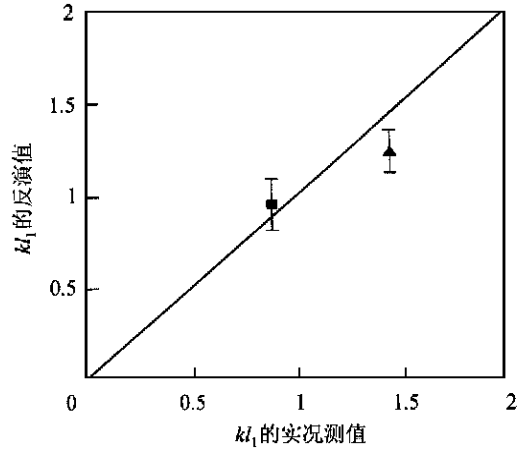


图 3  $kl_1$  的实测值与反演值的比较

Fig. 3 Comparison between retrieved  $kl_1$  and ground-truth data

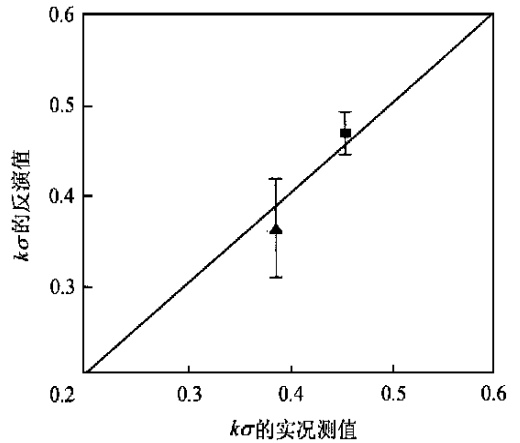


图 4  $k\sigma$  的实测值与反演值的比较

Fig. 4 Comparison between retrieved  $k\sigma$  and ground-truth data

图 5 是两块土壤含水量  $m_v$  的反演结果(BA3 有 8 月 14 日与 15 日两个不同的含水量)。土壤的含水量是通过反演得到的介电常数( $\epsilon'$ ,  $\epsilon''$ )计算到的。

图 6 是根据对 BA2 土壤所反演的平均粗糙度和平均湿度，采用双尺度模型所作的后向散射系数随入射角度变化的曲线，图中的离散点为实际测量值。

在以上反演过程中，由双尺度近似模型模拟计算的  $\sigma_{hh}^r$  对代价函数  $C$  的影响是明显的。如果待反演的地表实况能较好地满足双尺度模型，则反演结果较好。否则，必须加大代价函数的值，而产生较大的反演误差。当然，双尺度近似的模型不是粗糙面散射的唯一模型，因此，只要用作模拟  $\sigma_{hh}^r$  的模型合

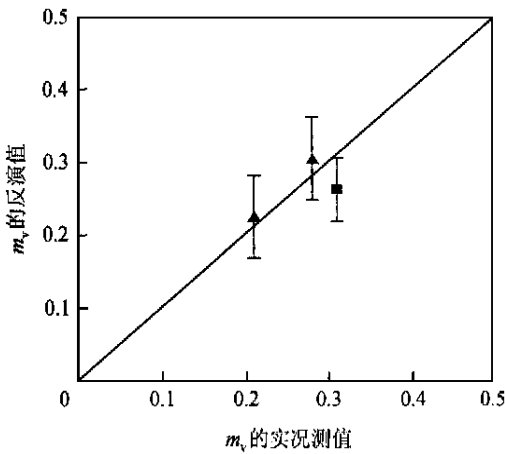


图 5  $m_v$  的实测值与反演值的比较

Fig. 5 Comparison between retrieved  $m_v$  and ground-truth data

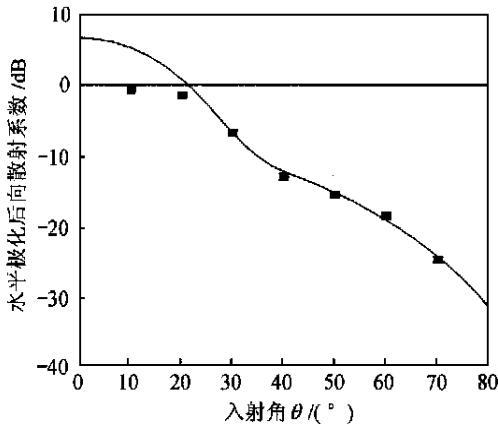


图 6 后向散射测量值与由反演参数产生的模拟结果

Fig. 6 Backscattering data and model simulation by using retrieved parameters

适,代价函数构造得好,该遗传算法的过程都会产生好的反演结果。

## 5 结论

本文提出将遗传算法应用于从遥感实测后向散

射系数来同时反演土壤湿度和地表粗糙度的新方法。将待反演参数编码形成基因,不同的参数组合形成不同的染色体,一定数量的染色体形成生物群体。通过双尺度近似的模拟计算构造代价函数,衡量反演结果的适应度,进行自然选择、克隆、杂交、变异等遗传算子操作,以寻求地表粗糙度和土壤湿度的最佳组合。本文根据车载遥感实测数据,对不同湿度和不同粗糙度的土壤进行了反演,得到的反演结果与地表实况的实际测值作了比较,结果是令人满意的。遗传算法的反演不用像 ANN 那样需要大量的数据去长时间训练网络,只要有多个角度的几个合适的实验数据,就可以进行多维参数的同时反演,这是遗传算法的一个明显优点。

## 参考文献 (References)

- [1] Huang X, Jin Y. Q. A Simple Method to Estimate the Soil Wetness and Surface Roughness from Active/Passive Data [J]. *Remote Sensing of Environment*, 1995, **53**(3): 212—214.
- [2] Jin Y. Q., Liu, C. Biomass Retrieval from High-Dimensional Active/Passive Remote Sensing Data by Artificial Neural Networks [J]. *International Journal of Remote Sensing*, 1997, **18**(4): 971—979.
- [3] Haupt R. L. Thinned Arrays Using Genetic Algorithm [J]. *IEEE Trans. on Anten. Propag.*, 1994, **42**(7): 993—999.
- [4] Sarabandi, K, Li, E. S. Characterization of Optimum Polarization for Multiple Target Discrimination Using Genetic Algorithm [J]. *IEEE Trans. on Anten. Propag.*, 1997, **45**(12): 1810—1817.
- [5] Holland J. H. Genetic Algorithm [J]. *Scientific American*, 1992: 66—72.
- [6] Goldberg D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning* [D]. Reading, MA: Addison-Wesley, 1989.
- [7] Haupt R. L. An Introduction to Genetic Algorithms for Electromagnetics [J]. *IEEE Anten. Propag. Magazine*, 1995, **37**(2): 7—15.
- [8] Jin Y. Q. *Electromagnetic Scattering Modelling for Quantitative Remote Sensing* [M]. Singapore: World Scientific, 1994.
- [9] Wegmuller U., Matzler C. *Active Passive Microwave Signature Catalogue (2-12 GHz)* [R]. University of Bern, Switzerland, 1993.

# A Genetic Algorithm to Simultaneously Retrieve Land Surface Roughness and Soil Moisture

WANG Yue-quan, JIN Ya-qiu

(Center for Wavescattering and Remote Sensing, Fudan University, Shanghai 200433, China)

**Abstract:** A novel genetic algorithm is developed for multi-parameter retrieval of land surface roughness and soil moisture. The theoretical model of two-scale rough surface is employed for computation of the cost function. Retrieved results by this genetic algorithm are well compared with ground-truth data. This study presents a new example of the genetic algorithm for multi-parameter retrieval in remote sensing.

**Key words:** genetic algorithm; roughness and wetness; retrieval