

用遗传算法优化估计Van Genuchten方程参数

许小健

合肥工业大学土木建筑工程学院, 安徽合肥(230009)

E-mail: Vividew@163.com

摘要: 描述土壤水分特征曲线的Van Genuchten方程是一个超定非线性复杂方程, 采用传统的方法对方程参数进行回归处理往往因为计算复杂和人为因素的影响, 而使估计结果带有较大的误差。基于自然选择和自然基因机制的遗传算法是当前处理一般非线性数学模型优化的一种新的优秀方法。为此, 采用实数编码多子种群遗传算法(PPGA)对方程的参数进行优化估计, 并基于Matlab环境编写了相应的计算程序。算例结果表明, PPGA能够很好地模拟实测数据, 且用PPGA优化估计Van Genuchten方程参数具有求解速度快、计算精度和自动化程度高、人为干扰因素小、通用性强等优点。因此, PPGA可作为计算Van Genuchten方程参数的一种新方法。

关键词: 遗传算法; Van Genuchten 方程; 参数优化

中图分类号: S152.7

1. 引言

土壤水分特征曲线是描述土壤含水量与吸力(基质势)之间的关系曲线。它反映了土壤水能量与土壤水含量的函数关系, 因此它是表示土壤基本水力特性的重要指标, 对研究土壤水滞留与运移有十分重要的作用^[1,2]。作为土壤水分滞留基础的土壤水分特征曲线问题, 国外学者先后提出了一些数学模型和计算方法, 其中Van Genuchten方程较为典型^[1-4], 因为其线型与实测数据曲线很相似, 而且参数意义明确。然而, Van Genuchten方程的函数表达式是一个超定非线性复杂方程, 含有的未知参数较多, 采用传统方法对方程参数进行处理, 不仅计算复杂, 通用性差, 而且人为因素的影响较大, 往往使预测结果带有较大的误差, 不易得到全局最优解。而基于自然选择和自然基因机制的遗传算法是当前处理一般非线性数学模型优化的一种新的优秀方法。因此, 为了提高Van Genuchten方程参数的估计精度, 本文提出用实数编码多子种群遗传算法(Real Coding Based Poly-Population Genetic Algorithm, 简称PPGA)对方程参数进行优化估计。基于MATLAB软件环境编写了PPGA的实现程序, 程序通用性强, 便于对数据进行修改。最后进行了算例分析。

2. 优化估计Van Genuchten方程参数的遗传算法

描述土壤水力特性的Van Genuchten方程表达式为[1]:

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{\left(1 + |\alpha h|^n\right)^m} \quad (1)$$

式中: θ 为土壤体积含水量($\text{cm}^3 \cdot \text{cm}^{-3}$); h 为土壤基质势(cm); θ_s 、 θ_r 分别为土壤饱和含水量($\text{cm}^3 \cdot \text{cm}^{-3}$)和残余含水量($\text{cm}^3 \cdot \text{cm}^{-3}$); α 、 n 、 m 为土壤水分曲线参数, $m = 1 - 1/n$, $0 < m < 1$ 。该方程的参数估计, 就是由实测数据序列 $\{(h_i, \theta(h_i))\}$ 确定 θ_s 、 θ_r 、 α 、 n 、 m 这5个参数的值。

不失一般性, 设 $F(C, X)$ 为一般非线性系统模型[5], 其中, $C = \{c_{ij} | j=1 \sim p\}$ 为模型 p 个待优化参数(优化变量), X 为模型 N 维输入向量, F 为模型 M 维输出向量, 也即 $F: R^N \rightarrow R^M$ 。根据 m 对模型输入、输出数据 $\{(X_i, Y_i) | i=1, 2, \dots, m\}$ 求优化问题:

$$\min f = \sum_{i=1}^m \|F(C, X_i) - Y_i\|^q \quad (2)$$

$$s.t. \quad a_j \leq c_j \leq b_j$$

式中： f 为优化准则函数； $[a_j, b_j]$ 为 c_j 的变化区间； $\| \cdot \|$ 为取范数， q 为实常数，当 $q=1$ 时为最小一乘准则，当 $q=2$ 时为最小二乘准则等等，可视实际建模要求而定。

考虑到实测数据序列 $\{(h_i, \theta(h_i))\}$ 样本容量较小，故采用稳健性好的误差绝对值之和最小准则构造优化准则函数：

$$\min f = \sum_{i=1}^m |\theta(h_i) - \theta(h_i)'| \quad (3)$$

$$s.t. \quad m = 1 - 1/n, n > 1$$

式中： $\theta(h_i)$ 为实测土壤体积含水量($\text{cm}^3 \cdot \text{cm}^{-3}$)； $\theta(h_i)'$ 为式(1)的计算值。当(3)式取最小值时对应的一组 θ_s 、 θ_r 、 α 、 n 、 m 即为所求最优参数估计值。这样，式(3)是一个非线性参数估计问题，由于参数 θ_s 、 θ_r 、 α 、 n 、 m 所处位置的不同，对优化准则函数式(3)计算值所做发挥作用也就不同。因此传统的方法较难处理该方程的参数估计问题。据笔者搜集到的文献，现有的估计方法大体上有：将最小二乘法和单纯形法相结合的解法[2]、借助最小二乘法利用Picard迭代方法[6]、单纯形调优法[7]、阻尼最小二乘法[8]等。其中，文献[6]的迭代方法虽然可以控制计算精度，但是如何构造具有物理意义且收敛的迭代公式和如何合理选取初值，这往往需要较高的数学理论知识。文献[2,7,8]的局限在于所选初值对拟合值有较大影响，需要试算才能获得较好的结果。因此，这些方法计算所得方程的参数估计值并不能保证是最优的。

基于自然选择和基因机制的标准遗传算法 (Simple Genetic Algorithm,简称SGA) [9,10]是在20世纪60年代末期到70年代初期，由美国Michigan大学的John Holland与其同事、学生们研究形成的一种从试图解释自然系统中生物的复杂适应过程入手，模拟生物进化的机制来构造人工系统模型的方法。它是当前处理一般非线性数学模型优化的一种新的优秀算法,它对模型是否线性、连续、可微等没有限制,也不受优化变量数目、约束条件的束缚,直接在优化准则函数(目标函数)的引导下进行全局自适应寻优,该方法直观、简便、通用、适应性强,至今仍是国内外常用的实施方案。但是，SGA在实际的应用中存在着计算量大，后期易早熟收敛等问题。而PPGA是利用多个群体同时进行进化寻优，在进化过程中，通过对优秀个体进行“移民”，既增加了群体的多样性，也增加了找到全局最优解的可能。该方法不仅有效地增加了避免早熟收敛的可能性，也增加了跳出局部最优趋向全局最优的可能性。为此，现用PPGA来优化估计Van Genuchten方程的参数。PPGA的实现步骤如下：

第1步： 编码。实数编码克服了二进制编码解码的换算占用计算机时间的问题以及表达精度的要求与计算量之间的矛盾。基于此，采用实数编码，利用线性变换：

$$x_j = a_j + y_j (b_j - a_j) \quad (4)$$

把初始区间为 $[a_j, b_j]$ 的第 $j(j=1, 2, \dots, p)$ 个优化变量 x_j 对应到 $[0, 1]$ 区间上的实数 y_j 。

第2步： 父代群体的初始化。生成 w (子种群数)个 z (每个子种群数包含的个体数)组 $[0, 1]$ 区间上的均匀随机数，每组 p (优化参数个数)个，即 $\{u_j\}$ ，把各 u_j 作为初始群体父代个体值 y_j 。经(4)式得到优化变量值，再经(3)式得到相应的目标函数值 $f(x_i)$ 。

第3步： 算出各个体的适应度值，并按线性排序进行评价。显然目标函数值 $f(x_i)$ 越小，

该个体适应能力越强，反之亦然。定义适应度函数为：

$$F(x_i) = 1 / (0.001 + f(x_i)) \quad (5)$$

式中的 0.001 是经验设置，为避免分母出现奇异的情况。

第4步：对每个子种群进行选择操作。从父代个体中以概率 p_i 按照比例选择方式选择第 i 个个体，共选择 n 组个体，按照轮盘赌选择：

$$p_i = F(x_i) / \sum_{i=1}^n F(x_i) \quad (6)$$

第5步：对每个子种群进行交叉操作。按下式进行随机线性重组产生：

$$\begin{cases} y'_j = u_1 y_{j1} + (1-u_1) y_{j2}, u_3 < 0.5 \\ y'_j = u_2 y_{j1} + (1-u_2) y_{j2}, u_3 \geq 0.5 \end{cases} \quad (7)$$

式中， u_1, u_2, u_3 都为[0,1]之间均匀分布随机数。

第6步：对每个子种群进行变异操作。变异操作决定了遗传算法的局部搜索能力。通过对每个父代个体的染色体上的基因以小概率 $p_m = 1 - p_i$ （变异概率）进行扰动，可以维持群体的多样性，防止早熟。

$$\begin{cases} y'_j = u(j), u_m < p_m(i) \\ y'_j = y_j, u_m \geq p_m(i) \end{cases} \quad (8)$$

式中， $u(j)(j=1,2,\dots,p)$ 和 u_m 均为随机数。

第7步：迁移进化。当迭代次数达到迁移世代后，把父代个体按目标函数值 $f(x_i)$ 从小到大排序。对排序后最前面 s 个个体(优秀个体)进行迁移操作，迁移方式为子种群之间任意的、无限制的迁移；否则,算法转入下一步。

第8步：收敛判别。算法运行达到预定进化次数或优个体的目标函数值小于某一设定值，也即达到设定的精度时，结束整个算法的运行，并把当前群体中最优个体作为 PPGA 求解方程参数最优估计值的结果，否则，转入第 3 步，重新进行评价、选择、交叉、变异和迁移操作。

笔者基于 Matlab 环境编写了模型的计算程序，针对不同组实测数据序列 $\{(h_i, \theta(h_i))\}$ ，只需修改程序中 h_i 和 $\theta(h_i)$ 值即可。程序通用性强，自动化程度高，且算法的控制参数设置简便。

3. 算例

为便于比较，本文以文献[2]的实测资料作为PPGA的计算原始数据。利用笔者编写的程序实验搜索5次，算法的控制参数可参照子种群个数取8个、每个子种群的个体数目取50个、每20代迁移个体、迁移率0.2、进化代数150代。考虑到 $m = 1 - 1/n$ ，故选择 θ_s 、 θ_r 、 α 、 m 这4个变量作为优化变量，其搜索范围根据经验可设置在[0,1]区间。程序所求Van Genuchten 方程参数估计值以及最优化准则函数值见表1，表1中也列举了文献[2]、文献[8]方法求得的方程参数值。由表1可见，3种方法计算的方程参数中， θ_s 、 α 值差别不大，但PPGA计算的 α 、 m 、 n 这3个参数值与文献[2]、文献[8]方法计算结果差别较大。PPGA与文献[2]模拟 θ - h 实测数据的效果比较见表2和图1，从表2的模拟数据和相对误差情况看，文献[2]对部分土壤体积含水量实测数据拟合偏小，拟合效果在图1中也很直观。PPGA优化估计方程参数的收敛过程见图2。

表1 不同方法计算的粉壤土Van Genuchten方程参数值
Table 1 Calculating Van Genuchten equation parameters in different methods

方法	$\theta_s / \text{cm}^3 \cdot \text{cm}^{-3}$	$\theta_r / \text{cm}^3 \cdot \text{cm}^{-3}$	α / cm^{-1}	m	n	最优化准则函数值
文献[2]	0.363	0.053	0.013	0.7642	4.241	0.108 234
文献[8]	0.363	0.053	0.014	0.7617	4.197	0.147 387
PPGA	0.363	0.057	0.014	0.6605	2.946	0.063 827

表2 PPGA与其他方法模拟粉壤土(脱湿) θ -h实测数据的比较
Table 2 Contrast data between measurement and different method simulations

基质势 h/cm	含水量 $\theta / \text{cm}^3 \cdot \text{cm}^{-3}$				相对误差/%		
	实测	文献[2]	文献[8]	PPGA	文献[2]	文献[8]	PPGA
5	0.363	0.363	0.363	0.363	-0.001	-0.001	-0.022
30	0.347	0.322	0.309	0.300	-7.304	-10.96	-13.46
70	0.279	0.262	0.242	0.254	-5.929	-13.37	-8.825
100	0.165	0.160	0.143	0.186	-3.273	-13.60	12.77
200	0.095	0.067	0.064	0.097	-29.66	-32.19	2.112
300	0.081	0.057	0.056	0.076	-29.93	-30.68	-6.703
500	0.069	0.054	0.054	0.064	-22.15	-22.30	-7.356
1000	0.059	0.053	0.053	0.059	-10.04	-10.06	-0.338
1300	0.054	0.053	0.053	0.058	-1.792	-1.798	7.557

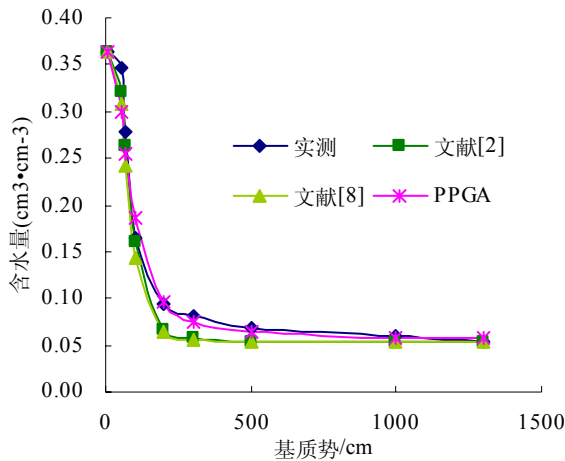


图1 不同方法模拟粉壤土 θ -h实测数据的效果比较
Fig.1 Data comparison between measurement and different method simulations

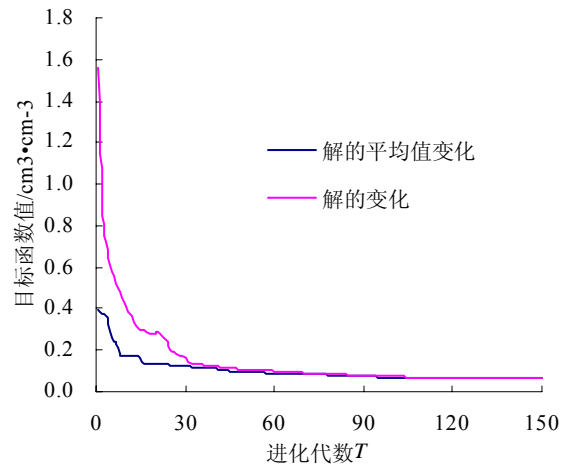


图2 PPGA求解Van Genuchten方程参数收敛过程
Fig.2 Convergence process of PPGA solving Van Genuchten equation parameters

为了进一步说明PPGA在优化估计Van Genuchten方程参数精度方面的优势和编写的程序的通用性强的特点,表3列出了PPGA优化估计不同土壤类型的Van Genuchten方程参数值,另外,表4、表5分别给出了PPGA与文献[2]方法模拟细砂、砾石土壤实测数据的对比情况(注:表中数据采用了四舍五入处理),可见本文PPGA法能够很准确地优化估计方程的参数值,且能够很好地模拟实测数据。

表3 PPGA优化估计的Van Genuchten方程参数值
Table3 Van Genuchten equation parameters estimated by PPGA

土的类型	$\theta_s / \text{cm}^3 \cdot \text{cm}^{-3}$	$\theta_r / \text{cm}^3 \cdot \text{cm}^{-3}$	α / cm^{-1}	m	n	最优化准则函数值
细砂脱湿	0.377	0.083	0.0148	0.7837	4.623	0.037 040
细砂吸湿	0.374	0.081	0.0152	0.8006	5.015	0.026 068
砾石脱湿	0.274	0.060	0.0239	0.6531	2.882	0.017 838
砾石吸湿	0.270	0.055	0.0243	0.6497	2.855	0.025 770

表4 PPGA与其他方法模拟 θ -h数据与实测数据的比较(一)
Table4 Contrast data between measurement and different method simulations

基质势h/cm		10	50	100	200	400	800	1000	1300	平均绝对 值误差/%	
含水量 θ /cm ³ ·cm ⁻³	实测	0.377	0.330	0.146	0.106	0.092	0.083	0.079	0.076	10.943	
	文献[2]	0.380	0.318	0.160	0.087	0.073	0.070	0.070	0.070		
	细砂脱湿	相对误差/%	0.796	-3.636	9.589	-17.925	-20.652	-15.663	-11.392	-7.895	5.000
	PPGA	0.377	0.330	0.146	0.089	0.083	0.083	0.083	0.083		
	相对误差/%	-0.009	0.016	0.067	-16.287	-9.274	0.046	5.085	9.219	8.937	
	实测	0.374	0.326	0.131	0.092	0.089	0.081	0.078	0.074		
	文献[2]	0.380	0.325	0.136	0.077	0.071	0.070	0.070	0.070	3.837	
	细砂吸湿	相对误差/%	1.604	-0.307	3.817	-16.304	-20.225	-13.580	-10.256		-5.045
PPGA	0.374	0.326	0.131	0.084	0.081	0.081	0.081	0.081	3.837		
相对误差/%	-0.005	-0.101	-0.208	-8.300	-8.754	0.016	3.853	9.462			

表5 PPGA与其他方法模拟 θ -h数据与实测数据的比较(二)
Table5 Contrast data between measurement and different method simulations

基质势h/cm		10	50	100	200	400	800	1000	1300	平均绝对 值误差/%	
含水量 θ /cm ³ ·cm ⁻³	实测	0.272	0.173	0.093	0.074	0.066	0.061	0.059	0.057	10.167	
	文献[2]	0.271	0.191	0.102	0.065	0.055	0.053	0.053	0.052		
	砾石脱湿	相对误差/%	-0.368	10.405	9.677	-12.667	-16.667	-13.115	-10.169	-8.772	3.051
	PPGA	0.272	0.172	0.099	0.071	0.063	0.061	0.061	0.060		
	相对误差/%	-0.084	-0.519	6.628	-3.927	-4.503	-0.292	2.610	5.841	5.425	
	实测	0.268	0.167	0.081	0.069	0.062	0.056	0.054	0.052		
	文献[2]	0.268	0.165	0.088	0.061	0.054	0.053	0.052	0.052	4.570	
	砾石吸湿	相对误差/%	0.000	-1.198	8.642	-11.594	-12.903	-5.357	-3.704		0.000
PPGA	0.268	0.167	0.094	0.066	0.058	0.056	0.056	0.056	4.570		
相对误差/%	-0.159	-0.241	16.561	-3.813	-6.190	-0.223	2.923	6.453			

4. 结语

PPGA克服了传统方法在分析过程中面临的复杂计算和人为因素的干扰,在可行解空间内同时从多个初始解开始搜索,防止搜索过程收敛于局部最优解,有较大可能求得全局最优解,并且对于待寻优的非线性模型的形式和参数估计个数等没有限制,因此可广泛应用于各类非线性模型的参数估计,具有广泛的适应性,是一种较理想的计算方法。

用PPGA对Van Genuchten方程参数进行优化求解,在一定程度上简化了Van Genuchten方程参数估计的困难,为求解Van Genuchten方程参数提供了一条新的途径。本文PPGA在优化估计Van Genuchten方程参数中的应用也验证了PPGA具有计算速度快、求解精度和自动化程度高、通用性强、适应性好等优点。

参考文献

- [1] Van Genuchten M. Th. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils[J]. Soil Sci. Soc. Am. J. 1980, 44 (5) : 892—898.
- [2] 王金生, 杨志峰, 陈家军, 等. 包气带土壤水分滞留特征研究[J]. 水利学报. 2000, (2):1-6.
- [3] 刘建立, 徐绍辉, 刘慧. 几种土壤累积粒径分布模型的对比研究[J]. 水科学进展. 2003, 14(5): 588-592.
- [4] 徐绍辉, 张佳宝, 刘建立, 等. 表征土壤水分持留曲线的几种模型的适应性研究[J]. 土壤学报. 2002, 39(4): 498-504.
- [5] 金菊良, 丁晶. 水资源系统工程[M]. 成都: 四川科学技术出版社, 2002. 102-103.
- [6] 徐绍辉, 张佳宝. 求土壤水力特征的一种迭代法[J]. 土壤学报. 2000, 37(2): 271-274.
- [7] 李春友, 任理, 李保国. 利用优化方法求算Van Genuchten 方程参数[J]. 水科学进展. 2001, 12(4): 473-478.
- [8] 马英杰, 虎胆·吐马尔拜, 沈冰. 利用阻尼最小二乘法求解 Van Genuchten 方程参数[J]. 农业工程学. 2005, 21(8): 179-180.
- [9] Holland J H. Genetic algorithms[J]. Scientific American, 1992, 9(7): 44-50.
- [10] 王小平, 曹立明. 遗传算法:理论、应用及软件实现[M]. 西安: 西安交通大学出版社, 2002. 10-38.

Estimating parameters by solving Van Genuchten equation using Genetic Algorithm

Xu Xiaojian

School of Civil Engineering; Hefei University of Technology, Hefei, China (230009)

Abstract

The Van Genuchten's Equation(VGE) of the curves which are used to describe the feature of soil water movement is super-set and nonlinear. Traditional method such as the least squares method calculating the VGE's parameters often results in a high margin of error because of complication and artificial factors. Genetic algorithms (GA) is a new outstanding evolutionary computation method to solve complex nonlinear model, it is based on natural selection and natural genetic mechanisms. Therefor, in this paper, real coding based poly-population genetic algorithm (PPGA) was applied on the VGE's parameters optimization. With the help of software Matlab environment, the solution program of PPGA was compiled, then some examples of PPGA estimating the VGE's parameters was given. The results show that PPGA solution can effectively fit the measured data, and PPGA is a high algorithm with many good properties such as high efficiency, high precision, rapid calculation speed, small artificial factors, etc. Consequently, PPGA can be used as a new method to calculate Van Genuchten equation parameters.

Keywords: genetic algorithm; Van Genuchten equation; parameter optimization

作者简介: 许小健, 男, 1984 年生, 硕士研究生, 主要研究方向是岩土工程数值计算方法。