

· 红外技术 ·

林地型背景植被温度估算模型

马栋良, 杨大峰, 刘国权, 胡峥峥

(61489部队, 河南 洛阳 471023)

摘要: 针对多数理论计算模型参数较多、实用性不强的缺点, 介绍了多元线性回归的原理及实现方法, 并对一组实验数据进行了分析, 得到了植被温度的简单模型, 并利用模型对植被温度进行了理论计算。通过计算结果与实测数值进行分析对比, 结果表明, 多元线性回归分析相对准确地估算出了植被的温度, 这为植被温度估算模型的建立提供了一个新的思路。

关键词: 温度; 多元线性回归; 模型

中图分类号: TN215; O434.3

文献标识码: A

文章编号: 1673-1255(2011)04-0032-03

Temperature Calculational Model of Woodland Background Vegetation

MA Dong-liang, YANG Da-feng, LIU Guo-quan, HU Zheng-zheng

(61489 Army Unit, Luoyang 471023, China)

Abstract: The most theoretical calculational models have the disadvantages of having too many parameters and little practicability, the theory and method of the multivariate linear regression are introduced. The temperature model of the vegetation is acquired by analyzing a set of testing data, and the model is calculated. By analyzing the calculated result and actual measured result, the result shows that the vegetation temperature is accurately estimated by the analysis of the multivariate linear regression, this offers a new idea to establish temperature calculational model of the vegetation.

Key words: temperature; multivariate linear regression; model

由斯蒂芬-波尔兹曼定律可知, 目标的热红外特性与目标的发射率和表面温度密切相关, 对于林地型背景来说, 植被的温度是其热红外特性很重要的方面。温度受到太阳辐射、天空辐射、气温、天气状况、相对湿度等许多因素有关, 它们之间的相互关系较复杂。当前的理论模型大多参数比较多, 具有很大的局限性, 不能直接应用在工程领域, 所以文章以试验数据为出发点, 通过对大量数据的分析, 得出他们之间的规律, 拟合出经验公式, 以便工程上应用。多元线性回归是处理大量数据一种重要的方法, 它通过对大量数据进行分析, 找出因变量与多个自变量之间关系。多元统计分析方法已经成功地应用于地质学的许多领域^[1]。文中采用多元线性回归模型的方法对一组实验数据进行分析, 对其在植被温度

估算模型中的应用进行了初步探讨。

1 多元线性回归简介

一组容量为 n 的实验数据

$$X = \begin{bmatrix} 1 & x_{i1} & \cdots & x_{im} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \vdots & x_{nm} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \beta = [\beta_0, \beta_1, \cdots, \beta_m]^T \quad (1)$$

式中, ε 为观测时的随机误差; ε_i 为第 i 次实验时的误差。则多元线性回归模型的矩阵形式为^[2]

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 I) \end{cases} \quad (2)$$

采用最小二乘法进行参数估计, 估计参数 β , 选

收稿日期: 2011-00-00

作者简介: 马栋良(1982-), 男, 河南南阳人, 硕士, 助理工程师, 从事红外技术研究。

取 β 的一个估计值 $\hat{\beta}$,使得误差的平方和达到最小,误差平方和为

$$Q(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) \quad (3)$$

求 β 使 $Q(\beta)$ 最小,即

$$\frac{dQ}{d\beta} = -2X^T(Y - X\beta) = 0 \quad (4)$$

求得

$$\hat{\beta} = (X^T X)^{-1} X^T Y \text{ 或 } \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m) \quad (5)$$

将 $\hat{\beta}$ 代入得到 y 的估计值得

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 + \dots + \hat{\beta}_m \cdot x_m \quad (6)$$

拟和残差即误差估计值为: $\hat{\varepsilon} = y - \hat{y}$;残差平方和为:

$$SS_{Re} = \sum_{i=1}^n (y_i - \hat{y}_i)^2; \text{总平方和为: } SS_T = \sum_{i=1}^n (y_i - \bar{y})^2; \text{回}$$

$$\text{归平方和为: } SS_R = SS_T - SS_{Re}; \text{相关系数为: } R^2 = \frac{SS_R}{SS_T}.$$

2 模型假设检验

实际问题中,在根据因变量与多个自变量的实际观测数据建立多元线性回归方程之前,因变量与多个自变量间的线性关系只是一种假设,建立方程之后,必须对因变量与多个自变量间的线性关系的假设进行显著性检验,通常用 F 检验方法。

$$\text{假设检验: } \begin{cases} H_0: \beta_j = 0 (j=1, \dots, m) \\ H_1: \beta_j \text{ 不全为0} \end{cases} \quad (7)$$

当 H_0 成立时,由概率统计知识可推导出: $SS_T / \sigma^2 \sim \chi^2(m)$, $SS_{Re} / \sigma^2 \sim \chi^2(n - m - 1)$ 。所以, $F = \frac{SS_T / m}{SS_{Re} / (n - m - 1)} \sim F(m, n - m - 1)$ 。

当 H_0 不成立时, F 的值有偏大的趋势,在给定显著水平 α 时,查 F 分布表得临界值 $F_\alpha(m, n - m - 1)$,计算 F 的观测值 F_0 。如果 $F_0 \leq F_\alpha(m, n - m - 1)$,接收 H_0 ,即因变量 y 与自变量 x_1, x_2, \dots, x_m 之间的线性关系不显著。如果 $F_0 > F_\alpha(m, n - m - 1)$,则拒绝 H_0 ,认为因变量 y 与自变量 x_1, x_2, \dots, x_m 之间的线性关系显著。

3 模型的建立

3.1 多元线性回归的Matlab实现

基本命令^[3-5]:

$b = \text{regress}(Y, X); [b, bint, r, rint, stats] = \text{regress}(Y, X, alpha);$

$$\text{输入: } Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{i1} & \cdots & x_{im} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \vdots & x_{nm} \end{bmatrix}, \text{alpha: 显}$$

著水平,一般取0.05。

输出: b :回归系数估计值, $bint$:回归系数置信区间, r :残差向量, $rint$:残差向量置信区间; $stats$: (s_1, s_2, s_3), s_1 :相关系数 R^2 , s_2 :模型检验 F 值; s_3 : F 分布随机变量取值大于 s_2 的概率。 $s_3 < alpha$ 时认为变量之间线性关系显著。

3.2 温度估算模型的建立

(1) 背景的选取

当前,林地型背景没有严格的定义,国防科技大学的曹义等在研究时采用的是樟树树冠^[6],解放军理工大学王友军等在研究时选取的是树木和草地^[7],为此,试验中选取一片树林占55%,草地占25%,裸露占20%的地域作为典型的林地型背景,试验选择5棵树和草地中3个地点为测量点,每个点测量5次,求平均值作为该测量点的温度值,然后用8个测量点的温度值代表该背景的植被温度,选择晴天、多云、阴雨天3种代表性的天气中进行测量。

(2) 试验数据

表1是2010年5月3个时间段内,在洛阳郊外某地植被温度及相应的气象条件进行的试验记录,3个时间段分别处于晴天、多云、阴雨天3种代表性的天气中。

植被的温度作为因变量 Y ,气温、湿度、照度作为自变量 X 。具体情况是: Y 为植被的温度; x_1 为气温值; x_2 为湿度值; x_3 为照度值; X 为 $[\text{ones}(60, 1) \ x_1 \ x_2 \ x_3]$;运行Matlab,结果如下:

$$\begin{aligned} b &= [20.508; 0.31877; -0.094123; 0.012537]; \\ bint &= [17.722 \ 23.295; 0.23987 \ 0.39768; -0.11557 \\ &\quad -0.072674; 0.0033921 \ 0.021681]; \\ stats &= [0.96496 \ 514.06 \ 0 \ 0.31523]; \end{aligned}$$

调用函数: $\text{rcoplot}(r, rint)$ 绘制残差向量及残差区间图。由图1可知,第18, 24, 38号数据偏差较大,残差区间不包含0点,将其去掉再进行运算,结果如下: $b = [21.167; 0.29197; -0.095416; 0.016379];$

$$\begin{aligned} bint &= [19.169 \ 23.166; 0.23558 \ 0.34836; -0.11089 \\ &\quad -0.07994; 0.009743 \ 0.023015]; \\ stats &= [0.98169 \ 947.23 \ 0 \ 0.15738]; \end{aligned}$$

表1 试验数据

时间	5月5日~5月6日				5月21日~5月22日				5月23日~5月24日			
	植被温度	气温	湿度	照度/($\times 10^4$)	植被温度	气温	湿度	照度/($\times 10^4$)	植被温度	气温	湿度	照度/($\times 10^4$)
10:00	20.2	25.8	0.39	48	23.3	24.1	45	2.22	22.2	22.7	54	0.125 6
11:00	26.4	26.1	0.41	63	24.4	25.4	43	2.26	22.6	23.2	51	0.132 6
12:00	27.3	27.8	0.40	65	26.5	26.8	32	2.585	23.6	23.8	48	0.153 2
13:00	29.5	27.5	0.34	63.5	27.9	27.7	28	2.465	24.0	24.7	47	0.144 4
14:00	25.4	29.2	0.34	58	27.6	28.1	25	2.55	24.6	25.6	48	0.143 4
15:00	25.8	27.9	0.35	47	25.7	26.2	33	2.298	24.2	23.9	46	0.137 8
16:00	23.3	24.7	0.44	35	24.9	24	38	1.8	22.5	23.1	56	0.135 6
17:00	23.2	24.6	0.48	10	23.7	23.8	43	1.23	21.7	22.8	60	0.127 6
18:00	23.1	22.3	0.50	2.2	22.5	21.8	53	0.854	21.2	21.9	65	0.065 3
19:00	21.3	21.3	0.56	0.001	22.4	21.1	48	0.001	20.6	21.5	67	0.001 2
20:00	22.1	18.3	0.73	0	21.4	18.2	50	0	19.8	19.7	73	0
21:00	22.1	18.2	0.62	0	21.4	18.1	51	0	19.6	17.4	76	0
23:00	20.6	17.7	0.69	0	20.7	17.9	56	0	18.9	16.6	81	0
1:00	19.2	15.5	0.70	0	18.2	15.7	76	0	18.5	15.5	82	0
3:00	18.3	13.9	0.77	0	18.0	15.2	73	0	18.2	15.2	79	0
5:00	18.1	13.1	0.79	0.027	17.5	14.6	77	0.001	18.0	14.8	79	0.000 8
6:00	19.4	14.9	0.79	1.17	18.4	15.4	70	0.045 3	19.8	16.9	70	0.034 5
7:00	19.9	21	0.61	24	19.3	19.3	57	0.654	20.8	21.3	71	0.067 8
8:00	22.4	22.2	0.59	33.5	21.8	21.8	60	1.568	21.5	21.7	65	0.105 4
9:00	24.1	25.2	0.48	41.5	22.4	23.7	54	2.1	22.4	22.5	61	0.114 7

变量之间存在显著的线性关系。所以,可以建立如下

$$Y = \begin{pmatrix} 21.167 \\ 0.291 97 \\ -0.095 416 \\ 0.016 379 \end{pmatrix} X \quad (8)$$

3.3 模型的试验验证

为了验证估算模型,在上述试验之后又进行了一次试验,表2是2010年5月28日10:00~23:00时间段内同一区域植被的温度、气象数据、模型估算温度值及温度估计值与实测值的差值。

表2 5月28日试验数据

时间	植被温度	气温	湿度	照度/($\times 10^4$)	模型估算值	温度差值
10:00	27.6	29.1	0.43	48	26.1	1.5
11:00	28.3	29.6	0.35	50.4	26.4	1.9
12:00	29.8	30.4	0.32	56.5	26.9	2.9
13:00	30.2	33.5	0.23	50.3	27.9	2.3
14:00	30	32.5	0.10	47	27.4	2.6
15:00	29.3	31.8	0.11	35.5	26.8	2.5
16:00	28.8	30.5	0.13	24	26.0	2.8
17:00	27.6	27.1	0.30	12.8	24.3	3.3
18:00	26.3	26.7	0.32	3.3	23.9	2.4
19:00	24.2	20.8	0.40	0.1	21.4	2.8
20:00	22.7	17.4	0.71	0	20.1	2.6
21:00	20.8	16.1	0.86	0	19.7	1.1
23:00	20.5	15.9	0.86	0	19.6	0.9

4 结论

由表2可知,计算值与实测值相差最小值为0.9℃, (下转第52页)

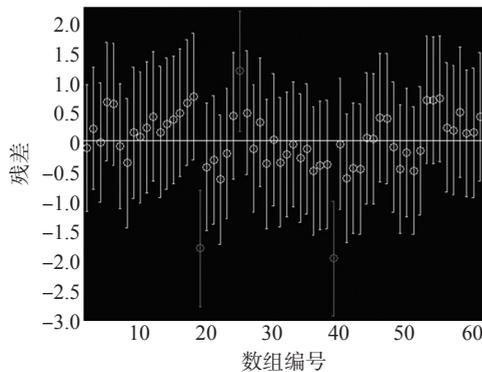


图1 残差向量与残差区间关系图

再次调用 $rcoplot(r, rint)$,生成图像如图2所示。由图2可知,数据没有异常,相关系数 R^2 为 0.981 69, 相对较大;该模型建立之初样本容量为 60, 中间剔除了 3 个数据, 因此样本容量 $n=57$, 自变量个数 $m=3$; $n-m-1=53$ 。查表计算^[8] $F_{0.05}(3, 53)=2.79 < F_0=947.23$; F 分布随机变量取值大于 S_2 的概率为 0, 显然 $0 < 0.05$, 由假设检验可知, 因变量与自

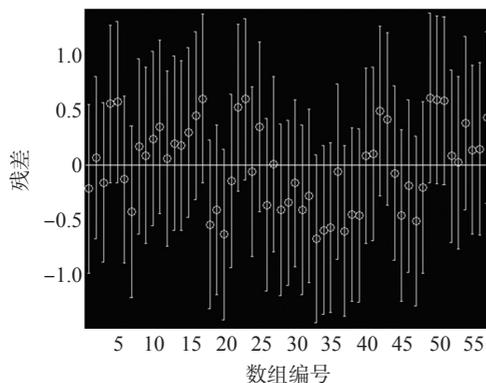


图2 剔除偏差点后残差向量与残差区间关系图

IQE 曲线,对实际电池生产或研发具有一定指导意义。

参考文献

- [1] Jianhua Z, Aihua W, Green M A. 24.5% efficiency silicon PERT cells on MCZ substrates and 24.7% efficiency PERL cells on FZ substrates[J]. Progress in Photov.: Research and Appl., 1999, 7: 471-474.
- [2] 阙端麟. 硅材料科学与技术[M]. 杭州: 浙江大学出版社, 1998.
- [3] 管绍茂, 王迅. 半导体表面钝化技术及应用[M]. 北京: 国防工业出版社, 1981.
- [4] Licht S, Maschhitz V, Lin B, et al. Chemical synthesis of battery grade super-iron barium and potassium Fe

(VI) ferrate compounds[J]. Power Source, 2001, 99: 7-14.

- [5] Vichai Meemongkolkiat, Kenta Nakayashiki, Dong Seop Kim. Factors limiting the formation of uniform and thick aluminum-back-surface field and its potential[J]. Journal of the Electrochemical Society, 2006, 153 (1): G53-H58.
- [6] J L Murray, A J McAlister. Bulletin of Alloy Phase Diagrams, 5.74(1984).
- [7] J D Alamo, J Eguren, A Luque. Operating limits of Al-alloyed high - low junctions for BSF solar cells[J]. Solid-State Electron., 1981, 24: 415-420.
- [8] Clugston D A, Basore P A. PC1D version 5: 32-bit solar cell modeling on personal computers[C]//26th IEEE Photovoltaic Specialists Conference, Anaheim CA, 1997: 207-210.

(上接第13页)

除。(2)衰减片具有不均匀性,各点衰减倍率不可能完全一致,经过衰减后光斑能量分布会有所改变,由此引入的误差将影响参数测量精度。(3)影响光斑质心测量精度的因素还包括 CCD 探测器的不均匀性,主要包括 CCD 的单个像元对激光响应的不均匀性。(4)CCD 摄像机存在位置量化误差,对于光束截面尺寸的测量精度由 CCD 器件的像素大小及像素密度决定。(5)CCD 摄像机存在灰度量误差,光强大小的测量精度由 CCD 器件的动态范围及图像采集系统的量化误差决定。

5 结束语

近年来,激光技术在工农业生产、卫生医疗、通

讯产业等领域得到了广泛应用,对激光参数的检测也提出了越来越高的要求。针对测试需求,设计开发了基于 CCD 的远场激光光斑测量系统,实现了远场激光光斑的采集与参数分析,在外场测试时获得了良好的效果。

参考文献

- [1] 张红娜,王祁. 图像测量技术及其应用[J]. 电测与仪表, 2003(40): 19-22.
- [2] 胡林亭. CCD 测量激光光斑方法研究[J]. 激光技术, 2001, 25(2): 154-157.
- [3] 刘兆辉,严彦文,许路铁. 未制导炮弹远距激光光斑监视系统[J]. 激光与红外, 2006(2): 88-91.
- [4] 庞长富,刘榴娣. CCD 摄像机用于测量中存在的问题及解决方法[J]. 光学技术, 1996(2): 5-8.

(上接第34页)

最大为 3.3°C。估算模型的计算结果还是可信的。同时模型的使用也是方便的,因为对于某特定地域的气温、湿度和照度可以由当地的气象部门获知,这样就很方便地估算出了当地植被的温度。但是必须经过长年累月的数据积累和处理才能得到相对精确的估算模型,以上分析只提供了一种建立估算模型的新思路。

参考文献

- [1] 刘昌蓉,何政伟,黄爽兵. 多元回归分析在地质灾害危险性评价中的应用[J]. 环保科技, 2008, 14(1): 31-33.

- [2] 张小蒂. 应用回归分析[M]. 杭州: 浙江大学出版社, 1991: 161-209.
- [3] 陈永胜. 多元线性回归建模以及 MATLAB 和 SPSS 求解[J]. 绥化学院学报, 2007, 27 (6): 166-168.
- [4] 王振友,陈莉娥. 多元线性回归统计预测模型的应用[J]. 统计与决策, 2008(5): 46-47.
- [5] 曾建军,李世航,王永国,等. MATLAB 语言与数学建模[M]. 合肥: 安徽大学出版社, 2005.
- [6] 曹义,才鸿年,程海峰,等. 林地型伪装网红外发射率取值分析[J]. 激光与红外, 2008(7): 671-675.
- [7] 王友军,胡江华,吕绪良,等. 林地型背景热红外伪装的温差阈值试验研究[J]. 红外技术, 2008(6): 361-364.
- [8] 盛骤,谢式千,潘承毅. 概率论与数理统计[M]. 北京: 高等教育出版社, 2001.