

文章编号: 1005-0523(2010)03-0027-06

# 回归分析在界限含水率试验数据处理中的应用

赵秀绍<sup>1</sup>, 莫林利<sup>2</sup>

(1. 华东交通大学 铁路环境振动与噪声教育部工程研究中心, 江西 南昌 330013; 2. 华东交通大学 软件学院, 江西 南昌 330013)

**摘要:**常用的界限含水率数据处理方法中, 作图法不够精确, 公式法仅能同时处理三组试验数据。针对以上特点, 提出了线性回归计算界限含水率的优化算法。该算法可以充分利用平行试验的多组数据, 使误差达到最小。重点阐述了用最小二乘法计算界限含水率的基本原理, 给出了程序设计流程图, 并用 Visual Basic 语言实现了这种算法, 最后利用典型工程实例分析了该算法的优缺点。

**关键词:**界限含水率; 回归分析; 优化算法; Visual Basic

**中图分类号:** TP311+TU411

**文献标识码:** A

界限含水率试验所得结果液限  $w_L$  和塑限  $w_p$  是确定细粒土的承载能力、对细粒土进行填料分组的重要指标, 所以准确地确定土的液塑限指标对工程建设具有重要的意义。

界限含水率试验测定完成后, 目前常用的数据处理方法有作图法和公式法。作图法受人为因素影响, 估读数据时一般会参照常规坐标, 造成读数错误, 其次每次试验都靠手工完成, 不能实现数据处理自动化。近年来许多学者提出了一些新的算法, 例如: 殷春娟提出了用 Excel 表格处理界限含水率的方法<sup>[1]</sup>, 彭意<sup>[2]</sup>, 冯朝辉<sup>[3]</sup>, 张传邦<sup>[4]</sup>提出了公式法计算界限含水率的方法, 高盟<sup>[5]</sup>提出了用样条函数求解的方法, 这些方法大大的提高了计算的效率与精度。以上方法能通过塑限差法 ( $\Delta w_p < 2\%$ ) 控制试验的数据偏差与精度, 但仅能处理三组数据, 部分试验数据被迫舍去而得不到充分利用, 往往导致试验要制样 5~9 次进行测试, 而有效的仅有 3 次。在前人工作方法的基础上, 提出了基于线性回归的界限含水率计算方法, 充分利用了各数据点对试验结果的贡献, 用 Visual Basic 语言实现了这种算法, 并在实践中检验了该算法的合理与优越性。

## 1 作图法计算界限含水率

根据文献[6]可得: 在液塑限联合测定试验中, 锥尖入土深度  $h$  (mm) 与含水率  $w$  (%) 之间存在如式(1)之间的双对数线性关系:

$$\lg h = a \lg w + b \quad (1)$$

式中:  $a$  和  $b$  是由试验确定的常数, 根据以上原理, 界限含水率试验的作图法应采用双对数坐标系。例如《TB10102-2004 铁路工程土工试验规程》<sup>[7]</sup> 或土工试验方法标准<sup>[8]</sup>法, 以  $w$  为横坐标,  $h$  为纵坐标, 在双对数坐标纸上绘制, 三组数据(A、B、C)应连成一条直线。三点明显不在一条直线上时, 连接 AB 和 AC 并延长, 与  $h=2$  mm 直线有两个交点  $D_1$ 、 $D_2$ ,

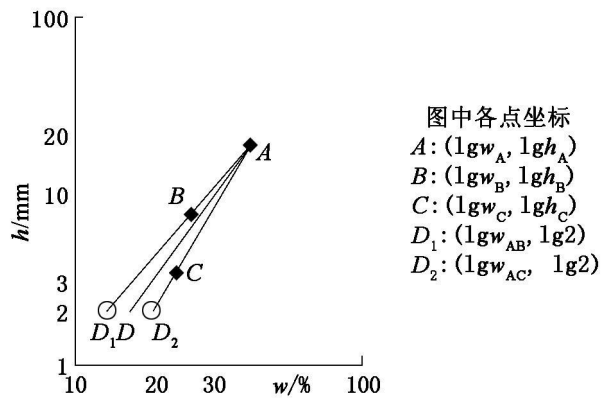


图 1 作图法原理图

收稿日期: 2010-01-12

基金项目: 江西省青年科学基金项目(GJJ09503), 华东交通大学博士科研启动基金项目(01306022), 华东交通大学科研

基金(09RJ05)  
中国知网 <https://www.cnki.net>

作者简介: 赵秀绍(1978-), 男, 博士, 副教授, 研究方向: 计算机在岩土工程中的应用。

两个交点横坐标( $w_{AC} - w_{AB}$ )之差称为塑限差,当  $\Delta w_p < 2\%$  时,取两个交点的中点  $D$ ,则  $AD$  就是表达式(1)的线性关系所绘制的直线。当  $\Delta w_p \geq 2\%$  时说明试验差距过大,应重做试验(见图1)。

作图法理论上可以计算和控制试验的精度,但实际操作起来相当困难,因为  $\Delta w_p \geq 2\%$  在双对数坐标上很难估读准确。

作图法的缺点:①受作图精度的影响;②受人为因素的影响;③作图法要求三个点之间的距离应尽量拉开,例如规程<sup>[7]</sup>规定, $C$ 点的范围在3~5 mm, $B$ 点范围在9~11 mm, $A$ 点范围在16~18 mm。但实际测定时,可能多次测试均不在以上范围内而需重新配水测定,造成试验数据浪费与人力资源浪费。

## 2 公式法计算界限含水率

依据作图法的原理(见图1),分别求出  $AB$  与  $AC$  直线的斜率  $K_{AB}$  和  $K_{AC}$ ,则  $K = (K_{AB} + K_{AC})/2$  可近似代替直线  $AD$  的斜率,利用式(2)和式(3)来求土样的液限  $w_L$  和塑限  $w_p$ <sup>[2-4]</sup>:

$$w_L = 10^{\frac{\lg(w_A) - \lg(K)}{\lg(h_A) - \lg(2)}} \times (1 - \lg(h_A)) + \lg(w_A) \quad (2)$$

$$w_p = 10^{\frac{\lg(w_A) - \lg(K)}{\lg(h_A) - \lg(2)}} \times (\lg(2) - \lg(h_A)) + \lg(w_A) \quad (3)$$

式中: $w_p$ 为塑限,锥尖入土2 mm对应的含水率; $w_L$ 为液限,锥尖入土10 mm对应的含水率。

公式法的优点:(1)公式法通过  $\Delta w_p < 2\%$  控制试验的精度;(2)公式法为计算机程序设计提供了依据,可以编写程序在计算机上实现自动化计算,此处不再详述。

公式法的缺点:(1)公式法是作图法的公式化,仅能依据三组数据来计算,而不能充分利用平行实验数据及超过了规范  $h$  范围的数据;(2)以含水率最大点  $A$  点作准确点是人为假设的,如果试验时  $A$  点的数据出现偏差,从而造成整个测定数据是错误的。

## 3 线性回归计算界限含水率方法原理

目前常用的线性回归方法为最小二乘法,它是德国数学家高斯在1794年解决行星轨道预测问题时首先提出的,其主导思想为保证拟合误差的平方差最小。

由式(1)可知含水率  $w$  与锥尖入土深度  $h$  取了对数后线性相关,所以可以用线性函数进行拟合。可令  $y = \lg w$ ,  $x = \lg h$ , 则式(1)转换为拟合方程

$$y = ax + b \quad (4)$$

设方程(4)的系数  $a, b$  已经确定,则对每一组试验自变量  $x_k(\lg w)$ , 都可以计算出一个  $y'_k(\lg h)$

$$y'_k = ax_k + b \quad (5)$$

式中: $k$ 是第  $k$  次试验; $y'_k$ 是  $y_k$  的计算值,  $y'_k$  与  $y_k$  之间的差值称为残差,用  $e_k$  表示。

$$e_k = y_k - y'_k = y_k - (ax_k + b) \quad (6)$$

显然,  $e_k$  的大小可衡量被确定的系数  $a, b$  的好坏,反过来,好的系数确定应遵循使  $e_k$  最小这样一个原则。最小二乘法就是使  $e_k$  的平方和达到最小,即

$$Q = \sum e_k^2 = \sum (y_k - y'_k)^2 \quad (7)$$

很显然,  $Q$  是系数  $a, b$  的函数,要使  $Q$  达到最小,则有  $Q$  对  $a$  和  $b$  的偏导数均为0,从而可得关于  $a, b$  的线性方程组<sup>[9]</sup>

$$\begin{cases} S_{11}a + S_{12}b = S_{1y} \\ S_{21}a + S_{22}b = S_{2y} \end{cases} \quad (8)$$

式中: $S_{11}$ 为试验总次数,  $S_{12} = S_{21} = \sum x_k$ ,  $S_{22} = \sum x_k^2$ ,  $S_{1y} = \sum y_k$ ,  $S_{2y} = \sum y_k x_k$ 。

解方程组(8)可得系数  $a, b$ 。

根据式(1), (4)可得

$$w = 10^{(\lg h - b)/a} \tag{9}$$

分别把  $h = 2 \text{ mm}$  和  $h = 10 \text{ mm}$  代入式(9)可得土样的液塑限。

线性函数最小二乘法拟合是靠相关系数来评价拟合好坏的, 相关系数用下式表示

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{10}$$

对于一般的线性回归, 相关系数的平方  $R^2$  小于 0.9 时数据是不可信的, 需重新试验, 或舍去偏离较大的点。

根据以上原理, 线性回归计算土样的界限含水率流程图如图 2 所示。

根据图 2, 本算法采用 Visual Basic 语言实现, 其实现的关键代码如下:

```
Public SubHuiGui (X () As Double, Y () As Double, A As Double, B As Double, R2 As Double)
DimS11 As Integer
DimS12 As Double, S22 As Double, S1y As Double, S2y As Double, delta As Double
S12 = 0; S22 = 0; S1y = 0; S2y = 0
S11 = UBound(X) + 1
For i = 0 To UBound(X) '计算方程组(8)的系数
    S12 = bb + X(i)
    S22 = AA + X(i) * X(i)
    S1y = dd + Y(i)
    S2y = cc + X(i) * Y(i)
Next
delta = S22 * S11 - S12 * S12
If Abs(delta) < 0.000 000 000 1 Then
    MsgBox "error"
Else
    A = (S2y * S11 - S12 * S1y) / delta '解方程组
    B = (S22 * S1y - S2y * S12) / delta
End If
R2 = 0 '计算相关系数
Dim Averx #, AverY #, SumX #, SumY #, Sum1 #, Sum2 #, Sum3 #
For i = 0 To UBound(X)
    SumX = SumX + X(i)
    SumY = SumY + Y(i)
Next
Averx = SumX / (1 + UBound(X))
AverY = SumY / (1 + UBound(X))
For i = 0 To UBound(X)
    Sum1 = Sum1 + (X(i) - Averx) * (Y(i) - AverY)
    Sum2 = Sum2 + (X(i) - Averx) ^ 2
    Sum3 = Sum3 + (Y(i) - AverY) ^ 2
Next i
R2 = Sum1 ^ 2 / (Sum2 * Sum3)
```

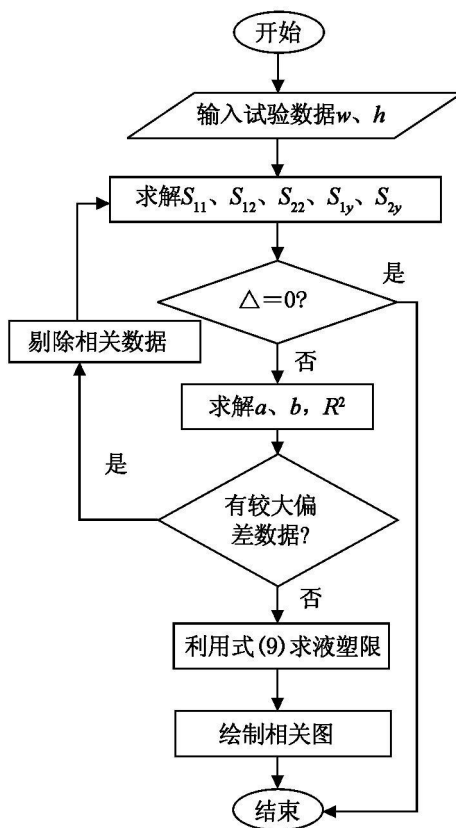


图 2 最小二乘法求解流程图

上述拟合函数调用如下:

```
Private Sub Command2_Click()
Dim I As Integer, J As Integer
Dim w() As Double, H() As Double '数据输入采用 VB 中的 Excel 控件表格输入
Dim MaxCow As Integer
MaxCow = Spreadsheet1.ActiveSheet.UsedRange.Columns.Count '确定试验数据组数
ReDim H(Maxrow - 2)
ReDim w(Maxrow - 2)
For I = 2 To Maxrow
H(I - 2) = Lg(Spreadsheet1.Cells(I, 2))
w(I - 2) = Lg(Spreadsheet1.Cells(I, 3)) '含水率和锥尖入土深度先对数后再进行回归分析
Next
Dim A As Double, B As Double, R2 As Double
HuiGui H(), w(), A, B, R2 '回归分析调用关键词句, A, B, R2为返回值
Dim wL10 As Double, wp2 As Double, Ip As Double
wp2 = 10 ^ (A * Lg(2) + b) '求解塑限 wp
wL10 = 10 ^ (A * Lg(10) + b) '求解 10 mm 液限 wL
Ip = wL10 - wp2 '求解液性指数 Ip
End Sub
```

#### 4 线性回归计算方法的工程应用与优缺点分析

本文选定南昌孔目湖粉土、樟树丁家山粉质粘土及抚州粘土 3 个典型实例,表 1 是 3 个工程土样界限含水率测定的数据,其中南昌孔目湖 104, 105, 106 对应的为平行试验数据。按照作图法和公式法的要求,试验超过 3 组数据的其余试验不能参与运算,同时这三组数据也不在规范范围内,都应舍去。如果采用了上述的最小二乘法来计算界限含水率,所有数据均能充分利用,其计算结果见表 2,其计算机绘制的拟合直线如图 3 所示。

从以上的工程实例中得出,基于最小二乘法有下列优点:(1) 数据不受三组试验组数的限制,平行试验数据可以得到充分利用;(2) 可以充分利用非规范范围内的数据,例如,公式法和作图法受规范锥尖入土深度范围的限制(3~5 mm, 9~11 mm, 16~18 mm),实践表明  $h$  只要在 3~20 mm 的锥尖入土深度范围,都能较好的满足方程(1),也即满足了最小二乘法的应用基础。

在工程应用中也发现当试验数据中某点(试验操作错误)偏离直线较大,已经超过了规范中的要求,但相关系数  $R^2$  仍大于 0.9,即无法完全通过相关系数来控制试验精度。其克服方法采用试验过程控制,即每一级锥尖入土深度测定两次,两次测定值之差应小于 0.7 mm<sup>[10]</sup>,否则直接重新试验,避免处理数据时发现问题返工重做试验。

表 2 表明,采用了上述试验方法后,塑限差最大值(由最大锥尖入土深度数据与其它数据组合成三组计算)都在 2%之内,达到了规范精度要求。

表 1 试验数据分析表

| 南昌孔目湖 |        |        | 樟树丁家山取土场 |        |        | 抚州仙临山取土场 |        |        |
|-------|--------|--------|----------|--------|--------|----------|--------|--------|
| 试验编号  | $h/mm$ | $w/\%$ | 试验编号     | $h/mm$ | $w/\%$ | 试验编号     | $h/mm$ | $w/\%$ |
| 101   | 4.775  | 20.426 | 201      | 4.450  | 32.795 | 301      | 5.000  | 44.128 |
| 102   | 8.820  | 23.832 | 202      | 8.000  | 38.288 | 302      | 7.860  | 50.005 |
| 103   | 19.360 | 31.389 | 203      | 10.700 | 41.321 | 303      | 10.430 | 55.149 |
| 104   | 5.675  | 21.126 | 204      | 14.450 | 44.244 | 304      | 16.500 | 64.629 |
| 105   | 11.920 | 26.032 | 205      | 19.200 | 48.430 |          |        |        |
| 106   | 15.060 | 28.489 |          |        |        |          |        |        |

表2 试验结果

| 计算指标                      | 南昌孔目湖   | 樟树丁家山取土场 | 抚州仙临山取土场 |
|---------------------------|---------|----------|----------|
| $w_p/\%$                  | 15.399  | 26.566   | 32.594   |
| $w_L/\%$                  | 25.155  | 40.546   | 54.646   |
| $I_p$                     | 9.756   | 13.980   | 22.052   |
| $R^2$                     | 0.990 7 | 0.998 5  | 0.996 5  |
| $\text{Max}\Delta w_p/\%$ | 1.467   | 0.291    | 1.772    |

注:  $\text{Max}\Delta w_p(\%)$  为最大塑限差, 规范控制指标, 由锥尖入土深度最大点也其余点连线求出塑限差距最大值。

采用以上回归算法, 可以很方便的设计作图程序, 其中南昌孔目湖与樟树丁家山土样界限含水率制图如图3与图4所示, 其相关系数  $R^2$  分别为 0.990 7, 0.998 5, 可见试验点相关显著, 其最大塑限差分别为 1.467%, 0.291%, 完全满足规范规定的  $<2\%$  的要求, 说明不在规范范围内的锥尖入土深度数据(南昌孔目湖  $h=8.820, 19.360$  mm; 樟树丁家山取土场  $h=8.000, 14.450, 19.200$  mm)可以使用。

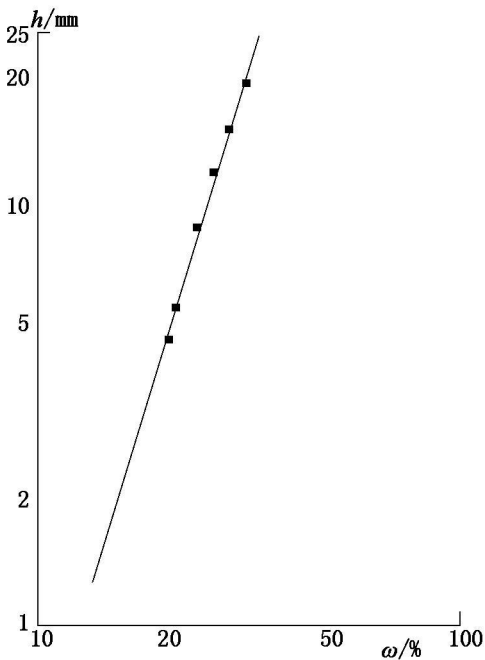


图3 有平行试验孔目湖界限含水率求解

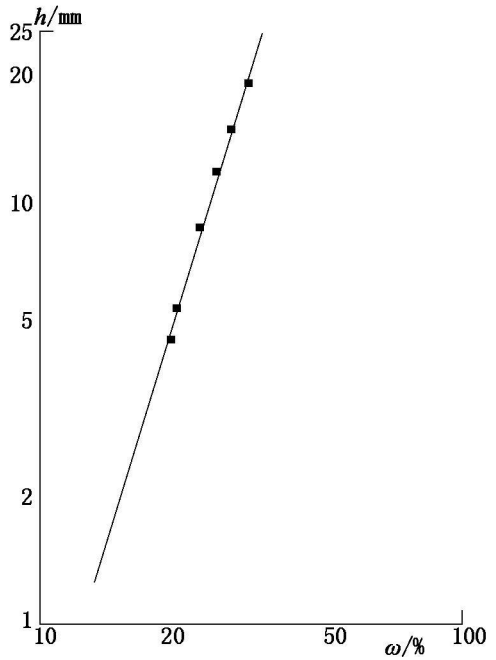


图4 樟树丁家山界限含水率求解

## 5 结论

本文详细介绍基于回归分析的界限含水率处理方法的原理与实现方法, 大大的方便了界限含水率的计算, 通过工程实践分析, 可以得出如下结论:

(1) 最小二乘法克服了作图法受人为因素、作图精度的影响, 且可以使数据处理过程自动化;

(2) 最小二乘法可以充分利于平行试验数据及规程范围之外的数据, 数据组数可以任意, 避免试验数据浪费及反复试验的繁琐, 且统计样本数据越多, 越能反映土样界限含水率实际情况;

(3) 最小二乘法可以使拟合后残差平方和达到最小, 克服了公式法把锥尖入土深度最大点  $A$  点(见图1)假设为准确点的不足, 实际上  $A$ 、 $B$ 、 $C$  点在试验中的位置是均等的。

## 参考文献:

[1] 殷春娟, 崔永高, 郑大龙. 液塑限联合测定法的若干问题探讨[J]. 上海地质, 2007, 27(1): 64-66.

[2] 彭意, 李铀, 彭强. 提高液塑限联合测定试验精度的方法[J]. 土工基础, 2007, 21(4): 81-84.

- [3] 冯朝辉, 彭意. 联合测定法中液塑限含水率的求取[J]. 湖南理工学院学报: 自然科学版, 2006, 19(2): 85-87.
- [4] 张传邦. 计算法求土的液限及塑限界限含水量[J]. 路基工程, 2006, 23(3): 28-29.
- [5] 高盟, 张远芳. 样条函数在液塑限联合试验成果确定中的应用[J]. 水利与建筑工程学报. 2005, 3(3): 8-10.
- [6] 南京水利科学研究所土工研究所. 土工试验技术手册[M]. 北京: 人民交通出版社, 2003.
- [7] 中华人民共和国国家标准. 铁路工程土工试验规程(TB 10102-2004)[S]. 北京: 中国铁道出版社, 2004.
- [8] 中华人民共和国国家标准. 土工试验方法标准(GB/T 50123-1999)[S]. 北京: 中国建筑工业出版社, 1999.
- [9] 吴若峰, 乐之伟, 陆文聪. 计算机在化学化工中的应用[M]. 上海: 上海大学出版社, 2000: 3-8.
- [10] 赵秀绍, 莫林利. 界限含水量试验数据计算方法及试验优化研究[J]. 路基工程, 2010, 27(2): 29-31.

## Application of Regression Analysis in Data Processing of Critical Moisture Content Experiment

Zhao Xiushao<sup>1</sup>, Mo Linli<sup>2</sup>

(1. Engineering Research Center of Railway Environmental Vibration and Noise, Ministry of Education; 2. Software School, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** Among common data processing methods of critical moisture content, mapping method is not very precise, and formula method can only deal with three sets of experimental data. To solve these shortcomings, an optimization algorithm of critical moisture content by linear regression method is proposed. The algorithm can take full advantage of data in the parallel tests to minimize the errors. The basic principle of the least squares method and the programming flow chart to calculate critical moisture content are proposed, and the algorithm has been realized with Visual Basic language. Finally, the advantages and disadvantages of the algorithm are analyzed by using several classical engineering project examples.

**Key words:** critical moisture content; regression analysis; optimization algorithm; Visual Basic

(责任编辑 王全金)