

基于 Excel 平台土壤含水量多元回归预测模型

邹文安, 姜波, 张薇

(吉林省墒情监测中心, 吉林 长春 130033)

摘要:土壤含水量是表述土壤干湿程度,反映旱情最直接的重要指标。土壤含水量预测能够反映未来某一时段农牧业旱情发展趋势,为开展旱情预警、各级领导和政府部门指挥抗旱减灾提供决策性依据。以降水、蒸发、风速等实测信息源为影响因子,以 Excel 为技术平台,创建了土壤含水量多元回归预测模型。该预测模型创建方法简单易行,便于改造和移植,有进一步推广价值。

关键词:Excel 平台;土壤含水量;多元回归;预测模型

中图分类号:S152

文献标识码:A

文章编号:1000-0852(2015)02-0065-05

1 引言

Excel 是微软公司的办公软件 Microsoft Office 的组件之一。具有数据处理、统计分析、简单的数据库管理和绘制图表等功能。尤其可利用 Excel 工作表,通过内置的统计函数、逻辑函数和数学函数等编程,能够实现多影响因子条件下的水文要素预测及模型创建等。

土壤含水量与自然地理、气象水文、土壤质地、下垫面、作物种植等多种因素有关,只有充分考虑各类因素对旱情的影响,才能建立起科学可靠的旱情预测模型。为此,土壤含水量预测应充分考虑降水、蒸发、温度、风速等因素影响^[1-2],选择科学合理的预测方法和预测技术平台,实现多影响因子条件下的土壤含水量预测。

回归分析预测法是根据系统元素之间的相互作用、相互影响,系统中事物发展变化的过程是许多其他因素共同作用的综合结果。这些因素与变量之间常常存在统计依赖关系,但却不能用确定的函数形式加以描述。可以依据大量的观测、统计数据,利用控制论的“黑箱”模型原理,找出这些因素之间的统计规律,并选用合适的数学方程式加以描述。常用的回归分析预测法主要有一元线性回归、多元线性回归等^[2]。

本文以 Excel 为技术平台,采用降水、蒸发、风速

等实测信息源为影响因子,选择多元回归为预测方法,创建土壤含水量多元回归预测模型。

2 关于多元回归预测

多元回归预测是数学模型预测法中的线性模型,是对模型中未知量进行次数划分,通常包含两个或两个以上自变量,且能够通过线性代数方程表达出未知量与多个自变量之间关系,称为多元回归线性预测模型。

根据多元回归分析的原理,建立起预报要素多元线性回归方程:

$$Y = b + m_1X_1 + m_2X_2 + m_3X_3 \cdots + m_nX_n \quad (1)$$

式中:Y 为预报要素,未知量;b 为常数项; X_i 为预报要素影响因子,自变量; m_1, m_2, \dots, m_n 即为回归系数(权重系数); $i=1, 2, \dots, n$ 为影响因子个数。

多元线性回归方程中的回归系数确定尤为关键。通常情况下,采用已知预报要素样本和影响因子样本进行耦合,反演出回归系数。参数输出结果见表 1。表 1 中: m_1, m_2, \dots, m_n 为回归系数;b 为常数项; $Se_b, Se_1, Se_2, \dots, Se_n$ 分别为回归系数的标准差; R^2 为复相关系数 R 的平方; Se_y 为 y 估计值的标准差;F 为方差比(F 统计量);df 为自由度; SS_{reg} 为回归平方和; SS_{resid} 为残差平方和。

在回归系数分析过程中,会衍生出复相关系数、

表1 回归系数分析计算参数输出形式样表

Table 1 The form sample of the output parameters for regression coefficients

m_n	m_{n-1}	……	m_2	m_1	b
Se_n	Se_{n-1}	……	Se_2	Se_1	Se_b
R^2	Se_y				
F	df				
SS_{reg}	SS_{resid}				

标准差等参数成果,这些参数可用于预测模型检验,评价预测模型和回归变量(影响因子)的显著性效果。检验方法分为 F 统计值(回归效果显著性检验)和 t 临界值(影响因子作用显著性检验)两种。

预测模型经显著性检验合格后,可进行预报要素预测。具体方法是:将各影响因子 X_i (预见期内的统计量)和回归预测模型分析出来的 b, m_1, m_2, \dots, m_n 回归系数代入回归方程(1),预测模型会自动输出计算值,该计算值即为预测要素预测成果。

3 土壤含水量多元回归预测模型

3.1 预报站简介

双辽墒情站位于吉林省东辽河流域,1981年建站,隶属吉林省水文水资源局。监测项目包括:土壤含水量(2001年开始监测)、降水、蒸发、温度、湿度、风向、风速等,该站处于吉林省西部平原半干旱区域,监测资料和地域分布均具有代表性。

3.2 数据样本

选用双辽站 2001~2012 年系列同步水文气象实测资料。其中,2001~2010 年系列实测资料作为回归系数分析计算(耦合)数据样本集,2011~2012 年系列实测资料作为检验预测数据样本集。

经分析,蒸发、降水、平均气温、风速与土壤含水量相关显著。为此,选用前期土壤含水量、旬累计蒸发量、旬累计降水量、旬累计日平均气温、旬累计日平均风速作为输入样本(影响因子);土壤含水量作为输出样本

(预测因子),预见期为 10 天。样本对应关系见表 2。

3.3 预测模型的建立

第一步:建立样本集

资料来源于 2001 年至 2010 年 3 月~6 月的各旬实测数据成果,按照表 2 样本对应关系,建立输出样本集(Y)、输入样本集(X_i)。包括实时土壤含水量,以及各旬累计的降水量、蒸发量、温度、风速样本系列共 50 组耦合样本数据,用于回归参数分析。

同理,选择 2011~2012 年系列 3 月~6 月 10 组实测数据作为检验预测样本,用于预测精度分析。

第二步:建立预测模板

预测模板包括耦合样本区、参数返回区、检验预测区三个部分,在一个 Sheet 工作表创建。

耦合样本区是数据源存放区域,采用 Excel 单元格赋值的方法提供。该区域以 Microsoft Excel 为技术平台,将已建好的输出样本集(Y)、输入样本集(X_i),复制到一张空白 Excel 中的 Sheet 工作表,生成单元格区域(如图 1 单元格 B7~G56)即为创建的耦合样本数据区。

参数返回区是耦合样本集数据回归计算后返回的回归系数区域(如图 1 中单元格 A73:F77)。创建方法为:选定任意单元格(如图 1 单元格 A73),输入回归函数分析计算指令,即“=LINEST(G7:G56,B7:F56,TRUE,TRUE)”(其中 G7:G56 为输出样本集,B7:F56 为输入样本集)后,单元格(A73)立即就会返回一个数据(如-0.0506);再以该单元格为基点,选定参数返回区(区域列数等于输入样本个数+1,行数为 5,如单元格 A73:F77),接着按 F2,再按组合键 Ctrl+Shift+Enter,于是参数返回区(单元格 A73:F77)就会返回一组数据,该组数据就是由输出样本集、输入样本集数据耦合产生的回归系数,见图 1、表 3。

回归系数返回成果值(见表 3)表明:双辽站多影响因子条件下的土壤含水量总体相关系数为 $R=0.7281$,表明输出、输入样本总体相关显著;前期土湿(前期

表2 输入/输出样本对应关系表

Table 2 The input/output samples

输出样本 Y		输入样本 X_i			
土壤含水量	土壤含水量	降水量	蒸发量	温度	风速
1 日实测	上月 21 日实测	上月下旬累计	上月下旬累计	上月下旬累计	上月下旬累计
11 日实测	1 日实测	上旬累计	上旬累计	上旬累计	上旬累计
21 日实测	11 日实测	中旬累计	中旬累计	中旬累计	中旬累计

基于Excel技术平台的多元回归预测模版												
预测站: 双辽												
预测内容: 土壤含水量; 影响因子: 前期土壤含水量、降水、蒸发、温度、风速												
耦合样本区												
5	序号	自变量样本 X_i					因变量样本 Y_i					
6		土壤含水量 (上期实测)	蒸发	降水	气温	风速	土壤含水量 (实测值)					
7	1	11.4	28.5	16.0	97.6	35.8	10.3					
8	2	12.7	48.7	1.0	187.0	19.8	10.4					
9	3	10.4	63.6	0.0	203.4	19.5	9.4					
54	48	10.2	33.4	12.1	187.9	7.3	11.2					
55	49	11.2	36.7	10.1	181.3	24.4	7.6					
56	50	7.6	16.6	0.0	130.7	24.1	7.1					
检验预测区												
58	序号	自变量样本 X_i					因变量样本 Y_i	土壤含水量 (实测)	误差	相对误差 (%)	合格否	
59		土壤含水量 (上期实测)	蒸发	降水	气温	风速	土壤含水量 (预测值)					
60	1	9.0	33.6	15.6	135.0	8.7	12.2	9.8	2.4	24.7	√	
61	2	9.8	43.4	23.2	192.8	12.6	13.5	10.3	3.2	31.4		
62	3	10.3	33.4	15.5	187.9	7.3	11.5	9.8	1.8	18.8	√	
68	9	10.1	55.5	12.7	217.9	32.3	10.7	11.4	-0.7	-5.9	√	
69	10	9.9	38.6	14.1	166.7	26.2	11.1	10.4	0.7	6.8	√	
70	检验样本合格率										90%	
参数返回区												
72	风速系数	气温系数	降水系数	蒸发系数	前期土壤系数	常数	返回统计参数					
73	-0.0506	-0.0221	0.2240	0.0785	0.4376	5.5442	回归系数及常数					
74	0.0421	0.0077	0.0226	0.0360	0.0864	1.4115	标准差值					
75	0.7281	2.0179	#N/A	#N/A	#N/A	#N/A	总体相关系数, 范围在0到1之间, 越接近1越好, 表示相关关系密切; 预测数据集的标准误差					
76	23.6	44	#N/A	#N/A	#N/A	#N/A	统计观察值; 自由度					
77	479.7	179.2	#N/A	#N/A	#N/A	#N/A	回归平方和、残差平方和					

图1 Excel 条件下多元回归预测模型界面

Fig.1 The multiple regression prediction model interface in Excel condition

表3 双辽站回归参数耦合后返回成果

Table 3 The coupled regression parameters and return results for the Shuangliao station

风速系数	气温系数	降水系数	蒸发系数	前期土湿系数	常数	参数说明
-0.0506	-0.0221	0.2240	0.0785	0.4376	5.5442	回归系数 m_i 及常数 b
0.0421	0.0077	0.0226	0.0360	0.0864	1.4115	标准差值 Se_i
0.7281	2.0179					总体相关系数 R ; 预测数据标准
23.6	44					统计观察值 F ; 自由度 df
479.7	179.2					回归平方和、残差平方和

土壤含水量)回归系数较大且正相关,表明土壤含水量受前期土湿影响较大;降水与土壤含水量为正相关,影响程度比其他影响因子显著;风速、气温与土壤含水量负相关且回归系数较小,说明其影响程度

相对较弱。

检验预测区(见图1中的B60~K69)是建立起土壤含水量与各影响因子多元回归线性关系,以及对土壤含水量预测精度的检验与评价。根据多元回归分析

原理可知,土壤含水量(预测值)可通过线性代数方程表达出来,即可利用 Excel 中的单元格(代表未知量、自变量、回归系数)公式链接形式表示。根据公式(1)和回归系数分析成果,那么双辽站土壤含水量实时预测多元回归线性方程为: $W_{\text{土壤含水量预测}} = 5.5442 - 0.0506 X_{\text{风速}} - 0.0221 X_{\text{气温}} + 0.2240 X_{\text{降水}} + 0.0785 X_{\text{蒸发}} + 0.4376 W_{\text{前期土湿}}$,对应 Excel 单元格(某一组)表达形式为:G60=F73+E73×B60+D73×C60+C73×D60+B73×E60+A73×F60,以此类推。

至此,完成了多影响因子土壤含水量多元回归预测模版的建立。

第三步:模型显著性检验

F 统计值检验:

返回成果值(见表 3)显示:样本统计观察值 $F=23.6$,自由度 $df=44$,样本集数据组数 $n=50$,于是自由度 $V_1=n-df-1=50-44-1=5$,自由度 $V_2=df=44$ 。假定显著性水平为 0.05,由 F 分布表可查得相应自由度的 F_a 临界值 $F_a=2.43$ 。很显然, $F=23.6$ 远大于 2.43,表明 F 统计值检验合理。

T 临界值检验:

T 检验观察值等于回归系数与标准差比值的绝对值。对应本次创建的土壤含水量多元回归预测模型 T 观察值(5 个影响因子就有 5 个观察值,即单元格 A73/A74、B73/B74……E73/E74)分别为 1.20、2.87、9.91、2.18 和 5.06;当自由度为 44、显著性水平为 0.05 的 t 临界值为 2.02,很显然,降水、前期土湿、气温、蒸发 T 观察值均大于 t 临界值,表明该 4 项自变量对土壤含

水量预测结果影响显著,可以用来预测土壤含水量;尽管风速 T 观察值(1.20)小于 t 临界值(2.02),考虑到风速对土壤含水量有一定影响,所以本次仍将风速作为自变量来预测土壤含水量。

3.4 预测精度检验

预测精度检验就是用创建的预测模型进行样本预测,来检验模型预测精度的质量。

本次预留了 2011~2012 年系列 10 组实测数据作为检验样本(不参加模型回归参数分析计算),将该样本拷贝到预测模板检验预测区的指定单元格,预测模型便能预测出土壤含水量(预测值)。通过比较土壤含水量预测值(计算值)与实际值误差,判断样本合格与否,统计样本集合格率等,实现预测模型检验与精度评价。

检验结果表明,10 组检验样本,有 9 组合格(预测值与实测值差值绝对值小于等于 3%,或预测值与实测值的相对误差小于等于 20%^[3-6]视为合格),合格率为 90%,见表 4、图 2。

4 模型应用

4.1 模型应用

本文采用土壤含水量、降水、蒸发、气温、风速 5 种影响因子多年系列资料作为输入样本信息源,以数据样本耦合形式反演预测模型参数,参数具有明显的区域属性。受其影响,该预测模型对有土壤含水量、降水、风速等水文气象观测资料的墒情站,代表性会很好;对没有上述观测资料但水文气象条件、土壤条件、

表4 双辽站多元回归预测模型预测精度检验成果表

Table 4 The multiple regression prediction model predictive accuracy test results for the Shuangliao station

序号	输入样本 X_i					输出样本 Y_i				
	土壤含水量 (上期)/%	蒸发 /mm	降水 /mm	气温 /°C	风速 /m·s ⁻¹	土壤含水量 (计算值)/%	土壤含水量 (实测) /%	差值	相对误差/%	合格否
1	9.0	33.6	15.6	135.0	8.7	12.2	9.8	2.4	24.7	√
2	9.8	43.4	23.2	192.8	12.6	13.5	10.3	3.2	31.4	×
3	10.3	33.4	15.5	187.9	7.3	11.6	9.8	1.8	18.8	√
4	9.8	36.7	26.6	181.3	24.4	13.4	11.6	1.8	15.4	√
5	11.6	16.6	0.0	130.7	24.1	7.8	10.8	-3.0	-27.8	√
6	9.4	27.5	20.0	111.2	36.1	12.0	10.1	1.9	18.6	√
7	10.1	47.2	11.5	181.8	43.7	10.0	9.4	0.7	7.0	√
8	9.4	53.8	1.8	161.5	46.8	8.3	10.1	-1.8	-17.6	√
9	10.1	55.5	12.7	217.9	32.3	10.7	11.4	-0.7	-5.9	√
10	9.9	38.6	14.1	166.7	26.2	11.1	10.4	0.7	6.8	√
合格率/%										90

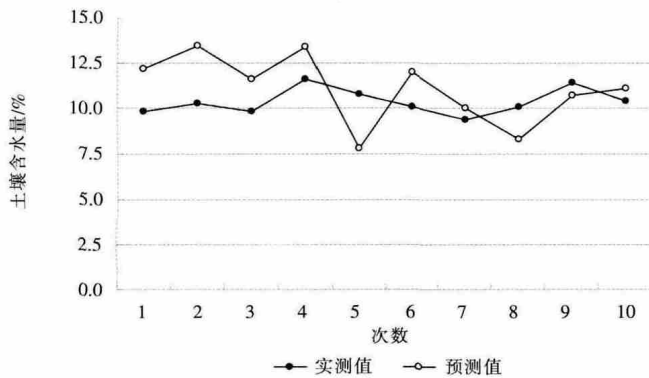


图2 双辽站多元回归预测模型检验过程图

Fig.2 The multiple regression prediction model checking process for the Shuangliao station

下垫面条件相似的邻近墒情站,也可以移用该模型进行土壤含水量预测。

本文选定了有土壤含水量、降水、风速等水文气象观测资料的双辽站进行多元回归预测模型数据样本耦合,反演回归系数,并正式进行土壤含水量生产预测。结果表明,一是模型考虑影响因素多。选用的前期土壤含水量、降水、蒸发、平均气温、风速均可作为土壤含水量预测影响因子,符合土壤含水量受多种水文气象因素共同影响的变化规律;二是模型预测精度高。利用该模型预测了2011年~2012年双辽站10组土壤含水量、降水、蒸发、平均气温、风速等实测数据样本,合格样本9组,合格率达90%;三是可用于区域预测。若选用预测站的水文气象、土壤质地、下垫面条件具有区域代表性,且那么该模型在一定程度上可用于区域预测^[7]。

4.2 模型拓展

为使该预测模型更具有操作性,实现输入样本(耦合输入样本、预测输入样本)自动读取,可创建一张空白Excel中的Sheet工作表,作为耦合输入样本、预测输入样本的数据输入表格,与创建的预测模型Sheet工作表指定单元格(一对一)定位链接,实现预测模型输入样本的自动读取功能。在此基础上,在影响因子(个数、组数)不变条件下(若影响因子有变化,可调整已创建的预测模型结构,使其与输入表格中的输入样本相同),可变换输入表格中的输入样本,耦合出该预测模型各类参数,经显著性检验合格后,可进行生产实际应用,实现预测模型移植、改变预测领域等功能,使预测模型适用性更广、操作性能更好。

5 结语

5.1 模型优缺点

优点:

(1)建模简单,使用方便,人机交互性能好,预测模型直观易懂;

(2)该模型基于Excel技术平台下建立的,其回归参数随着耦合样本集变化而自动改变,在影响因子数目相同时,创建的多元回归预测模型便于移植和改造;

(3)模型耦合样本、输出样本数据可视化,便于分析数据变化规律;

(4)模型具有相关系数计算、回归显著性判断(需通过人工干预)等功能;

(5)不需要任何商用软件或进行专门的模型软件开发,只要在常规电脑上基于Microsoft Office即可操作使用,既节约了模型开发成本,也方便了用户推广使用。

缺点:

(1)多因子共同影响预报对象时,模型有时只会识别敏感影响因子,且回归系数权重很大,对预测结果影响较大,这与实际往往不符;

(2)当影响因子为极端数值时,预报成果与实测值误差较大。

5.2 成果水平及应用

本文以Excel为技术平台,采用降水、蒸发、风速等实测信息源为影响因子,选择多元回归为预测方法,创建土壤含水量多元回归预测模型。该模型集数据处理、相关分析计算、参数分析率定、预测检验等多功能为一体的预测模型,为原始创新性的应用技术。预测模型直观易懂,创建简单,应用方便,且预测精度较高,可拓展到多影响因子条件下的水质预测、泥沙预测、枯季水资源、旱情预警,以及建筑成本预测、医学疑难病理诊断等。

参考文献:

- [1] 严昌荣,申慧娟.基于多元回归方法的土壤水分预测模型研究[J].湖北民族学院学报,2008,(9):241-245. (YAN Chang-rong, SHEN Hui-juan. Method based on multivariate regression prediction model of soil moisture [J]. Journal of Hubei Institute for Nationalities, 2008,(9):241-245. (in Chinese))
- [2] 吉林省墒情监测中心.吉林省墒情评价指标实验及早情分析技术研究[R].2012. (Soil Moisture Monitoring Center of Jilin Province. Jilin province soil moisture Indexes and analysis of drought exper-

(下转第89页)

西江水系的支流郁江和东江中下游支流西枝江的暴雨多由热带气旋造成。

(3)珠江流域较稳定的暴雨中心在长安~桂林、都安~迁江、桂平~昭平、英德~清远、河源~龙门等一带区域。发生流域性大洪水的暴雨其暴雨中心落在柳江的概率最大,其次是桂江,暴雨中心落在右江的概率最小。

(4)珠江流域洪水主要由暴雨形成的。锋面暴雨强度大,频次高,历时长,影响范围大,易造成流域性洪水,多发生在4月~6月前汛期。台风暴雨强度大,历时短,影响范围小,易形成局部大洪水,多发生在7月~9月后汛期。

参考文献:

[1] 水利部珠江水利委员会,《珠江续志》编纂委员会.珠江续志[M].北京:中国水利水电出版社,2009. (Pearl River Water Resources Commission of the Ministry of Water Resources, Codification Commis-

sion for Pearl River Conservancy Add History. Pearl River Conservancy Add History [M]. Beijing: China WaterPower Press, 2009. (in Chinese))

- [2] 薛纪善.1994年华南夏季特大暴雨研究[M].北京:气象出版社,1999. (XUE Jishan. Research on the Extraordinary Rainstorm of the South of China in the Summer of 1994 [M]. Beijing: China Meteorological Press,1999. (in Chinese))
- [3] 水利部水文局,水利部珠江水利委员会水文局.1998年珠江、闽江暴雨洪水[M].北京:中国水利水电出版社,2001.(Hydrology Bureau of Ministry of Water Resources, Hydrology Bureau of Pearl River Water Resources Commission of Ministry of Water Resources.The Storm Flood of Pearl River and Minjiang River in 1998 [M]. Beijing: China WaterPower Press, 2001. (in Chinese))
- [4] 水利部水文局,水利部珠江水利委员会水文局.2005年珠江暴雨洪水[M].北京:中国水利水电出版社,2007. (Hydrology Bureau of Ministry of Water Resources, Hydrology Bureau of Pearl River Water Resources Commission of Ministry of Water Resources.The Storm Flood of Pearl River in 2005 [M]. Beijing: China WaterPower Press, 2007. (in Chinese))

Analysis of Storm Weather System and Flood Characteristics in Pearl River Basin

YAO Zhangmin, DU Yong, ZHANG Lina

(Hydrology Bureau of Pearl River Water Resources Commission of the Ministry of Water Resources, Guangzhou 510611, China)

Abstract: This paper statistic analysed the storm weather systems, storm centers and occurrence time of the 9 model storm floods in Pearl River Basin, so as to seek the interrelation of the storm weather system, storm type, flood factor and occurrence time of different drainages of Pearl River Basin. This paper also summarized the essential characteristics of the storm floods in the Pearl River Basin, what can supply some reference value to flood control and disaster reduction in the Pearl River Basin.

Key words: storm; flood; weather system; spatial and temporal distribution; Pearl River Basin

(上接第 69 页)

imental technology research[R]. 2012.(in Chinese))

[3] SL 568-2012,土壤墒情评价指标 [S].(SL 568-2012,Evaluation Index of Soil Moisture Condition[S].(in Chinese))

[4] SL 364-2006,土壤墒情监测规范 [S]. (SL 364-2006,Technical Standard for Soil Moisture Monitoring[S].(in Chinese))

[5] SL 250-2000,水文情报预报规范[S]. (SL 250-2000,Hydrological Information and Forecasting Practices[S].(in Chinese))

[6] SL2013,土壤水分监测仪器通用技术条件(报批稿)[S]. (SL2013, General Technical Condition for Soil Moisture Monitoring Instruments(For Approval)[S].(in Chinese))

[7] 邹文安.基于Excel+VBA技术条件的土壤含水量预测模型[J].中国防汛抗旱,2013,(2):49-52.(ZOU Wenan. A soil moisture content prediction model based on Excel and VBA [J]. China Flood & Drought Mmanagement, 2013,(2):49-52.(in Chinese))

Multiple Regression Model for Soil Moisture Prediction on Excel-based Platform

ZOU Wenan, JIANG Bo, ZHANG Wei

(Moisture Monitoring Center of Jilin Province, Changchun 130033, China)

Abstract: Soil moisture content is expressed in soil moisture levels, reflecting the most direct and important indicator of drought. Soil moisture forecast to reflect the future development trend of a certain period of agricultural drought, for carrying out the drought early warning, command and leadership at all levels and departments to provide decision-making basis for drought mitigation. In this paper, precipitation, evaporation, wind and other sources of information as influencing factors measured to Excel as technology platforms, creating a multiple regression model for soil moisture prediction. The forecast model creation method is simple, easy to transform and transfer.

Key words: Excel platform; soil moisture; regression; prediction model