

田间土壤含水率的统计特性分析

李国芳 夏自强 郝振纯 蒋洪庚 陈海芳

(淮海大学水文水资源及环境学院 江苏 南京 210098)

摘要 :以实测土壤含水率资料为基础 ,通过假设检验方法系统地分析了田间土壤含水率的统计特性 .分析结果表明 ,田间土壤含水率为近似符合正态分布的随机变量 ,这一结论为田间土壤墒情监测误差的分析及控制提供了理论依据 .

关键词 土壤含水率 ;监测误差 ;随机 ;正态分布 ;统计检验

中图分类号 S152.7 文献标识码 :A 文章编号 :1000-1980(2002)01-0011-04

田间土壤墒情是农田耕作层土壤含水率的俗称 .田间土壤墒情的监测 ,作为非工程节水的配套技术 ,是农业用水和区域水资源管理的一项基础工作 ,对农田灌溉排水的合理实施、农作物的增产与节约用水、四水转化关系的分析和提高水资源利用率等均有重要的作用 .在田间土壤墒情监测中 ,通常将土壤含水率视为符合正态分布的随机变量 ,来分析监测误差和确定采样点数目^[1 2] .因此 ,利用广泛的实测资料检验土壤含水率的随机性和正态性是十分必要的 ,这直接关系到以此假设为基础的各种研究成果的正确性、可靠性 .

1 田间试验

为了研究土壤含水率在田间的变异规律及田间尺度的墒情监测方法 ,笔者在国家“九五”重点科技攻关项目“节水农业技术与示范”的示范区——山西省洪洞县霍泉灌区范围内 ,根据地貌、土壤质地和灌溉情况设置了李堡(1)、李堡(2)和虎头山三块试验地 .在 1997 和 1998 两年间进行了 10 次大规模的试验 ,获得了宝贵的实测资料 .

三块试验地的基本情况如下 :李堡(1)面积约 30 000 m² ,水浇地 ,土壤质地为粘壤土 ;李堡(2)面积约 5 000 m² ,水浇地 ,土壤质地为中壤土 ;虎头山面积约 5 000 m² ,旱地 ,土壤质地为轻壤土 .10 次田间试验中所选的土壤含水率观测方法有 TDR(时域反射仪)和 Moisture-probe 法 .根据实际情况 ,各次试验的垂向测点数及测点深度分三组 : (a)当垂向测点数为 6 时 ,TDR 法的测点深度分别为 0 ~ 10 ,10 ~ 20 ,20 ~ 30 ,40 ~ 50 ,60 ~ 70 ,80 ~ 90 cm ,Moisture-Probe 法的测点深度分别为 0 ~ 5 ,10 ~ 15 ,20 ~ 25 ,40 ~ 45 ,60 ~ 65 ,80 ~ 85 cm ; (b)当垂向测点数为 3 时 ,测点深度分别为 0 ~ 5 ,20 ~ 25 ,40 ~ 45 cm ; (c)当垂向测点数为 2 时 ,测点深度分别为 0 ~ 5 ,40 ~ 45 cm .各次试验的监测地点、监测时间、监测方法、农作物及监测点布设等基本情况见表 1 .

表 1 各次试验的基本情况

Table 1 Basic condition for each test

监测地点	测次	监测时间	监测方法	农作物	垂线数目	垂向测点数
李堡(1)	1	1997 - 06 - 03	I	小麦	20	6
	2	1997 - 11 - 27	II	小麦	52	2
	3	1998 - 07 - 14	II	玉米	50	6
	4	1998 - 11 - 16	II	小麦	84	3
李堡(2)	1	1997 - 06 - 04	I	玉米	12	6
	2	1998 - 07 - 15	II	玉米	20	6
	3	1998 - 11 - 17	II	小麦	35	3
虎头山	1	1997 - 06 - 05	I	小麦	12	6
	2	1997 - 07 - 17	I	玉米	15	6
	3	1998 - 07 - 17	II	玉米	26	6

注 : (a)监测方法 I 为 TDR 法 ,其探针长度为 10 cm ;监测方法 II 为 Moisture-probe 法 ,其探针长度为 5 cm . (b)各测点上探针均垂直埋设 .

2 田间土壤含水率的随机性检验

将同一测次同一深度所有测点的土壤含水率按相邻位置排成序列,则土壤含水率的非随机性主要表现为持续性、趋势性、周期性等,因此检验也应针对这几方面进行.

2.1 持续性检验

持续性检验采用序列自相关系数检验法.具体如下:

原假设 H_0 : 序列无持续性 统计量

$$t = r_1 \left(\frac{n-1-2}{1-r_1^2} \right)^{\frac{1}{2}} \tag{1}$$

其中

$$r_1 = \frac{\sum_{i=1}^{n-1} (x_i - \overline{x})(x_{i+1} - \overline{x_{i+1}})}{\sqrt{\sum_{i=1}^{n-1} (x_i - \overline{x})^2 \sum_{i=1}^{n-1} (x_{i+1} - \overline{x_{i+1}})^2}} \tag{2}$$

式中: n ——样本长度; x_i ——样本第 i 项的值; r_1 ——样本一阶自相关系数; $\overline{x_i} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i$; $\overline{x_{i+1}} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_{i+1}$.

当 n 较大时, t 渐近服从 $t_{(n-1-2)}$ 分布,分别取显著性水平 $\alpha = 10\% \ 5\% \ 2\% \ 1\%$, 求出 $t_{\alpha/2}$, 若 $-t_{\alpha/2} < t < t_{\alpha/2}$, 则接受 H_0 , 否则拒绝 H_0 .

样本一阶自相关系数反映了相邻测点土壤含水率间的相关程度,如果相邻测点土壤含水率的关系较为密切,则一阶自相关系数的绝对值较大, $|t|$ 值很大,必定拒绝原假设.因此,如果接受原假设,就说明各点田间土壤含水率较为独立.

2.2 趋势性检验

趋势性检验采用秩号-序号相关系数检验法,具体如下:

原假设 H_0 : 序列无趋势性 统计量

$$t = r_{s-o} \left(\frac{n-2}{1-r_{s-o}^2} \right)^{\frac{1}{2}} \tag{3}$$

其中

$$r_{s-o} = \frac{\sum_{i=1}^n (m_i - \overline{m})(i - \frac{n+1}{2})}{\sqrt{\sum_{i=1}^n (m_i - \overline{m})^2 \sum_{i=1}^n (i - \frac{n+1}{2})^2}} \tag{4}$$

式中: n ——样本长度; m_i ——样本中第 i 项的秩号; r_{s-o} ——样本秩号-序号相关系数; $\overline{m} = \frac{1}{n} \sum_{i=1}^n m_i$.

当 n 较大时, t 近似服从 $t_{(n-2)}$ 分布,分别取显著性水平 $\alpha = 10\% \ 5\% \ 2\% \ 1\%$, 求出 $t_{\alpha/2}$, 若 $-t_{\alpha/2} < t < t_{\alpha/2}$, 则接受 H_0 , 否则拒绝 H_0 .

序列从大到小排列后所得的秩号 m_i 和序号之间的相关系数反映了序列的趋势性.如果序列呈从大到小或从小到大的趋势,则相关系数的绝对值较大, $|t|$ 值很大,必定拒绝原假设.因此,如果接受原假设,就说明田间土壤含水率不存在一端比另一端明显干或明显湿的不均匀现象.

2.3 随机性检验

随机性检验采用转折点检验法.转折点是指序列中的极大点($x_i > x_{i-1}, x_i > x_{i+1}$)或极小点($x_i < x_{i-1}, x_i < x_{i+1}$).序列的转折点总数记为 P .

原假设 H_0 : 序列为纯随机序列 统计量: P .

当 n 较大时, P 近似服从 $N(a, b)$, 其中 $a = (2n-4)/3$, $b = (16n-22)/90$, 令 $\hat{u} = (P-a)/\sqrt{b}$, 则 $\hat{u}$ 近似服从标准正态分布 $N(0, 1)$. 分别取显著性水平 $\alpha = 10\% \ 5\% \ 2\% \ 1\%$, 求出 $u_{\alpha/2}$, 若 $-u_{\alpha/2} < \hat{u} < u_{\alpha/2}$, 则接受 H_0 , 否则拒绝 H_0 .

当按空间相邻次序排列的田间土壤含水率值呈一大一小交替出现的锯齿状时, P 值较大; 呈由大渐小、由小渐大的周期性时, P 值较小.两种情况下样本的随机性都较差, 当 P 接近于 a 时, 样本的随机性最好.

因假设检验对样本数目的要求,对前面所述各田块各测次所有测点数目大于 20 的平面(共 32 个)的土壤含水率资料进行假设检验.统计各单项检验接受原假设的平面占总平面数的百分比、各项检验均接受原假设的平面占总平面数的百分比(下称接受率),检验结果见表 2.

表 2 田间土壤含水率统计特性检验结果

显著性水平 α /%	接受率/%			
	持续性检验	趋势性检验	随机性检验	所有检验
10	68.8	71.9	87.5	56.3
5	68.8	84.4	100.0	65.8
2	71.9	87.5	100.0	68.8
1	81.3	90.6	100.0	81.3

同一序列是否接受原假设与所选的显著性水平 α 有关,显著性水平 α 越小,其接受域越大,越易接受假设.对于检验差异显著性和显著性水平的关系,一般规定:若显著性水平 $\alpha \geq 10\%$,所检验的统计量(例如趋势性检验中的 t)不越过置信临界值 $\pm t_{\alpha/2}$,则认为被检验序列与无趋势序列无真正的差异;若显著性水平 $10\% > \alpha \geq 5\%$,则认为被检验序列与无趋势序列可能有差异;若显著性水平 $5\% > \alpha \geq 1\%$,则认为被检验序列与无趋势序列有真正的差异^[3].由表 2 可见,当显著性水平 $\alpha = 10\%$ 时,各单项检验的接受率都不高,特别是全项检验的接受率更低.当显著性水平 $\alpha = 1\%$ 时,各单项检验及全项检验的接受率才较高.在各项检验中,持续性检验和趋势性检验的接受率较低,分析原因是由于有的田块某些测次可能因灌水、覆盖等的差异而略呈一侧干一侧湿的不均匀性.

总的来说,由于各种因素的影响,田间土壤含水率的随机性比其它水文变量如雨量、流量的随机性差,但在精度要求不高的情况下,将其视为随机变量是可以的.

3 田间土壤含水率的正态性检验

在客观实际中有许多随机变量,它们是由大量相互独立的随机因素综合影响所形成的,而其中的每一个别因素在总的影响中所起的作用都是微小的,这种随机变量往往近似地服从正态分布.田间土壤含水率是许多相互独立的随机因素共同作用的结果,因此,在研究田间土壤含水率的统计分布时,首先考察它是否服从正态分布.根据奥野忠一等人的研究,认为正态性检验方法中,偏度、峰度检验法是较为有效的方法之一^[4].

随机变量 X 的偏度和峰度指的是 X 的标准化变量 $(X - E(X))/\sqrt{D(X)}$ 的三阶中心矩和四阶中心矩:

$$\nu_1 = \frac{E((X - E(X))^3)}{(D(X))^{3/2}} \quad \nu_2 = \frac{E((X - E(X))^4)}{(D(X))^2} \tag{5}$$

这里 $E(\cdot)$ 表示数学期望函数, $D(\cdot)$ 表示方差函数.

偏度表示土壤含水率概率密度曲线的偏斜程度.偏度大于零,说明曲线呈正偏;偏度小于零,说明曲线呈负偏;偏度等于零,说明曲线对称.峰度表示土壤含水率概率密度曲线的陡缓程度.峰度越大,说明概率密度曲线越尖瘦;反之,说明概率密度曲线越矮胖.当随机变量服从正态分布时, $\nu_1 = 0$, $\nu_2 = 3$.

设 $x_1, x_2, \dots, x_n$ 是来自总体 X 的样本,则 ν_1, ν_2 的矩估计值 g_1, g_2 的计算公式如下:

$$g_1 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 / \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2} \quad g_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 / \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2 \tag{6}$$

若总体 X 为正态分布,则当 n 充分大时,近似地有

$$g_1 \sim N\left(0, \frac{6(n-2)}{(n+1)(n+3)}\right) \quad g_2 \sim N\left(3 - \frac{6}{n+1}, \frac{24n(n-2)(n-3)}{(n+1)(n+3)(n+5)}\right)$$

在“ X 为正态总体”的假设条件下,记

$$\sigma_1 = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \quad \sigma_2 = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)(n+3)(n+5)}} \tag{7}$$

$$\mu_2 = 3 - \frac{6}{n+1} \quad U_1 = g_1 / \sigma_1 \quad U_2 = (g_2 - \mu_2) / \sigma_2 \tag{8}$$

则 $U_1 \sim N(0, 1), U_2 \sim N(0, 1)$.

取显著性水平为 α 时,原假设的拒绝域为 $|U_1| \geq K_1$ 或 $|U_2| \geq K_2$.其中 K_1, K_2 取在原假设成立的条件下,使得 $P\{|U_1| > K_1\} = \alpha/2, P\{|U_2| > K_2\} = \alpha/2$ 的值.当 $\alpha = 10\%$ 时, $K_1 = K_2 = 1.96$.

计算出各地块各测次不同深度平面(共 50 个)上土壤含水率的 $|U_1|$ 和 $|U_2|$ 后,取显著性水平 $\alpha = 10\%$,则总共 50 组资料中有 42 组通过偏度、峰度检验,通过率达 84%,说明田间土壤含水率服从正态分布的假设是可以接受的.

4 应 用

前面已通过假设检验的方法论证了田间土壤含水率为服从正态分布的随机变量,基于这一结论,可以进行土壤含水率监测误差的分析及控制。

设土壤含水率 X 服从数学期望为 μ 、方差为 σ^2 的正态分布,则随机变量 $\xi = (\bar{X} - \mu) / \sqrt{\sigma^2/n}$ 服从标准正态分布,即 $\xi \sim N(0,1)$ 。 n 为样本容量, $\bar{X}$ 为 n 次样本的均值。由于总体方差 σ^2 经常是未知的,一般只能用样本方差 S^2 去估计总体方差 σ^2 。则由统计学原理,可构造一随机变量 $t = (\bar{X} - \mu) / \sqrt{S^2/n}$, t 服从自由度为 $n-1$ 的 t 分布。记土壤含水率的相对误差限 $\Delta = |\bar{X} - \mu| / \bar{X}$,则在显著性水平 α 下,有

$$\Delta \leq t_{\alpha/2(n-1)} \cdot S / (\bar{X} \cdot \sqrt{n}) = t_{\alpha/2(n-1)} \cdot C_v / \sqrt{n} \tag{9}$$

式中 C_v 为土壤含水率的变差系数。由此可见,土壤含水率测定的相对误差限与 C_v 和 $t_{\alpha/2(n-1)}$ 成正比,与样本容量 $n^{1/2}$ 成反比。而 $t_{\alpha/2(n-1)}$ 也随 n 的增大而减小,但仅当 n 较小时, $t_{\alpha/2(n-1)} \cdot n^{-1/2}$ 随 n 的增大而减小才显著,当 n 较大时,该项随 n 的增大而减小的速率就很小了。因此,土壤含水率测定的相对误差限随土壤含水率变差系数的增大而增大,随取样数目的增大而减小,但减小的速率越来越小。这说明虽然取样数目越多,测定误差越小,但一味增加取样数目只会增加监测工作量,对监测精度的提高作用不大。这就是合理的取样数目的确定问题。所谓合理的取样数目是使样本的均值有足够的准确度和较大的概率近似于真正的总体的均值所需的取样数。由式(9)可知,合理的取样数目

$$n = \frac{t_{\alpha/2(n-1)}^2 \cdot S^2}{\Delta^2 \bar{X}^2} = \frac{t_{\alpha/2(n-1)}^2 \cdot C_v^2}{\Delta^2} \tag{10}$$

因 n 的改变要引起 $t_{\alpha/2(n-1)}$ 的改变,故由式(10)求 n 值是一“试算”过程。
表3列出了按式(10)试算得到的显著性水平 $\alpha = 10\%$ 和 $\alpha = 5\%$,不同估值精度 Δ 时,对应不同变差系数 C_v 值的合理取样数目,可供实际应用时参考。

表 3 合理取样数目
Table 3 Necessary sampling number

α /%	Δ /%	C_v												
		0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
10	5	5	13	26	45	70	99	173	270					
	10	3	5	8	13	18	26	45	70	99	133	173	220	270
	15	2	3	5	7	10	13	21	32	45	61	78	99	122
	20	2	3	4	5	6	8	13	18	26	35	45	56	70
5	5	6	18	37	64	99	140	248	384					
	10	4	6	11	18	26	37	64	99	140	192	248	314	384
	15	3	4	6	10	13	18	30	45	64	86	111	140	172
	20	2	4	5	6	9	11	18	26	37	50	64	80	99

参考文献：

[1] 杨诗秀,雷志栋.田间土壤含水率的空间结构及取样数目确定[J].地理学报,1993,48(5):447~455.
[2] 陶士珩,王立祥,胡希远,等.土壤含水率测定的误差分析及控制[J].干旱地区农业研究,1997,15(2):84~88.
[3] 金光炎.水文水资源随机分析[M].北京:中国科学技术出版社,1992.83~86.
[4] 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,1989.222~225.

Statistical Characteristics Analysis of Soil Water Content

LI Guo-fang, XIA Zi-qiang, HAO Zhen-chun, JIANG Hong-geng, CHEN Hai-fang
(College of Water Resources and Environment, Hohai Univ., Nanjing 210098, China)

Abstract : Based on measured data, the statistical characteristics of soil water content are studied. Statistical tests show that the soil water content is a random variable, which approximately obeys the normal distribution. The conclusion provides a theoretical basis for the analysis and control of error in soil water content monitoring.

Key words : soil water content; monitoring error; random; normal distribution; statistical test