

# 干湿持续期随机模拟<sup>\* 1</sup>

王世耆 诸叶平 李世娟

(中国农业科学院农业信息研究所, 北京 100081)

## 摘 要

该文应用数据建模技术,实现干湿期随机建模。主要包括:利用历史气象资料,从中采集干湿期数据;应用实测数据,创建干湿期经验分布函数;应用 Monte Carlo 方法和经验分布参数,随机生成干湿期序列,通过和 Markov 链模型输出的对比分析,讨论生成序列的统计误差,测试结果显示,与两状态 Markov 链方法相比,所建模型性能更好。

关键词: 干湿持续期; 随机建模; 天气生成器

## 引 言

在农业、水文和环境等领域,随机天气模拟器(或生成器)有着广泛应用。它可生成长序列天气变量值,为有关数学模型提供必要的输入,还可满足其他一些数据需求,如:为缺少观测的地区插补逐日天气;在空间分析中,对每日天气数据做网格化处理以及构造气候变化情景。

天气生成器通常涉及多变量,如降水、辐射、最高气温和最低气温等,其中关键变量是降水。相应的降水模型普遍采用两状态一阶 Markov 链(简称 Mc)和日降水量 Gamma 分布函数的耦合结构<sup>[1-8]</sup>。在这类模型中,随机天气干湿状态由 Mc 决定,但用该模型生成的干湿期序列统计误差较大。而持续干旱和连绵霪雨,恰恰是影响我国经济民生、尤其是困扰农业生产的灾害性天气。

前人对干湿期的统计规律早有研究。人们曾假设逐日天气变化过程遵循柏努利测验,认为无论干日(或湿日)均对随后的干日(或湿日)无任何影响,或彼此随机不相干,以此推算干湿期的期望频数。Jorgensen 根据 Gold 和 Cochran 提供的方法,绘制了雨期延长概率图,用以估算雨期持续若干日后,至少再持续若干雨日的概率<sup>[9]</sup>。Steinhäuser<sup>[10]</sup>用简单的 Poisson 分布代表干湿期的频数分配,计算了各种长度的干期与湿期频数。Gabriel 等<sup>[11]</sup>认为干

湿期序列遵循简单的 Markov 过程,提出干湿期的概率分布。Longley<sup>[12]</sup>根据天气资料,用实验方法指出,干湿期序列长度与其累积频数间有对数关系。么枕生<sup>[13]</sup>认为,干日和湿日彼此并不独立,逐日天气变化不符合 Bernoulli 测试,而是有后效作用的链锁现象;并用 Polya 分布(亦称概率传染分布)计算了干湿期的期望频数。然而,已有的研究表明,所有尝试过的经典理论分布,似都不能满意地配合干期频数。

为满足农业研究对长序列逐日天气数据的需求,弥补现有天气生成器的不足,作者应用数据建模(modeling of data)<sup>[14]</sup>方法,构造了干湿期随机模型 DWS。与 Mc 不同,该模型使用的随机变量是干(或湿)期,而不是干(或湿)日;并直接应用经验分布来描述随机变量的统计特性,而不寻求分布函数的理论表述<sup>[15]</sup>。实践表明,该模型是可用的,且性能比 Mc 模型好。本文主要介绍建模步骤和测试结果。

## 1 干湿期观测值的生成

干湿期由干湿日组成。定义干(或湿)期,首先须定义湿日。通常将日降水量大于等于某个阈值(如 0.1 mm)的日子规定为湿日,小于阈值者为干日。持续为湿日的一段时期为湿期,持续为干日则为干期。因干、湿持续可能跨月和跨年,为实现按

\* 国家“863”计划课题(2006AA10Z220)、国家科技支撑课题(2006BAD10A06)及北京市自然科学基金课题(4042026)共同资助。  
2008-01-17 收到,2008-07-24 收到再改稿。

年、月统计,特别约定将干湿持续期的第 1 日所在年、月,定义为相应持续期所属年、月。图 1 说明某  $i$  月的统计约定。其中 D 表示干日, W 表示湿日。 $i$  月的最后 1 天为干日,由此延续到  $i+1$  月,又 10 个干日,这一干期长度共有 11 d,按约定,该干期应统计在  $i$  月的干期频数内,尽管这个干期有 10 d 是在  $i+1$  月。

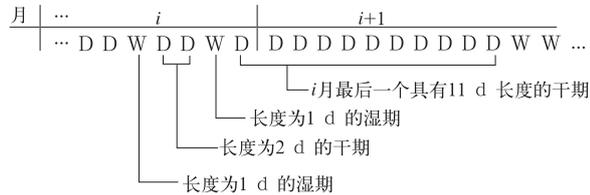


图 1 干湿期长度的统计约定  
Fig.1 Definition of dry and wet spells length monthly

因为在规范的气象记录中,并无显式的干、湿期统计数据,所以欲获取其观测值,只能从逐日降水资料中提取。本文设想逐日降水资料以某种数据库形式储存在计算机里。当从数据库中依次读取逐日天气记录时,遵循如下干、湿期观测值统计算法。

如读到 1 个干日(即该日降水量小于阈值)记录,则执行干期算法:

$$L_d = L_{d-} + 1, \text{ 若前条记录为干日。}$$

$$W_j = L_w, j = 0, 1, 2, \dots, n_w,$$

$$L_w = 0,$$

$$L_d = L_{d-} + 1, \text{ 若前条记录为湿日。}$$

如读到一个湿日(即该日降水量不小于阈值)记录,则执行湿期算法:

$$L_w = L_{w-} + 1, \text{ 若前条记录为湿日。}$$

$$D_i = L_d, i = 0, 1, 2, \dots, n_d,$$

$$L_d = 0,$$

$$L_w = L_{w-} + 1, \text{ 若前条记录为干日。}$$

其中,  $L_d, L_w$  分别是读出某天气记录时的干、湿持续日计数;  $L_{d-}, L_{w-}$  分别是读出前条记录时的干、湿持续日计数;  $D_i, W_j$  分别为干期第  $i$  个和湿期第  $j$  个观测值;  $n_d$  和  $n_w$  分别为干期和湿期观测值频数。

由于干湿期长度通常有显著的年变化,因此,至少按月提取观测值,并按月建立相应的经验分布,是必须的。

## 2 干湿期随机建模

### 2.1 经验分布函数及其统计

设某随机变量  $\xi$  的  $n$  次观测值为  $x_1, x_2, \dots, x_n$ 。

所谓经验分布函数  $F_n(x)$ , 系指由下式规定的函数:

$$F_n(x) = \begin{cases} 0, & x \leq x_1^*, \\ \frac{k}{n}, & x_k^* < x \leq x_{k+1}^*, \\ 1, & x > x_n^*. \end{cases} \quad (1)$$

式(1)中,  $x_1^* \leq x_2^* \leq \dots \leq x_n^*$  是按大小排序的观测值,  $x_k^*$  是第  $k$  个观测值。显然,  $F_n(x) = P^* \{ \xi < x \}$ , 其中  $P^*$  表示事件发生的频率。

根据 Гливленко 定理<sup>[16]</sup>, 当观测次数  $n$  增大时, 经验分布函数  $F_n(x)$  将接近理论分布函数  $F(x)$ , 即

$$P \left\{ \sup_{-\infty < x < \infty} |F_n(x) - F(x)| \rightarrow 0 \right\} = 1$$

显然, 在构建  $F_n(x)$  时, 总希望使随机变量观测样本足够大。为避免观测值频数为 0, 在建立干、湿期长度的经验分布函数时, 实际是按下式:

$$F_n(x) = \begin{cases} 0, & x \leq x_1^*, \\ \frac{n_k}{n}, & x_k^* < x \leq x_{k+1}^*, \\ 1, & x > x_k^*. \end{cases} \quad (2)$$

式(2)中,  $k$  是将  $x_1^*, x_2^*, \dots, x_n^*$  进行分组的组序,  $n_k$  表示第  $k$  组内的观测值频数。

考察干湿期长度观测值, 不难发现, 较短的干湿期常有较大的频数, 而较长的干湿期, 频数常常较小。因此, 在实践中, 试验了等距和不等距两种分组方案。

根据分组统计, 可得各组端值  $(x_i^*, x_{i+1}^*)$  和干湿期累积频率值  $F_i (i=1, 2, \dots, k)$ 。它们将被映射成计算机内一张累积频率分配表, 或模拟参数表, 将其绘制成图, 即得经验分布函数曲线。

### 2.2 干湿期随机模拟

根据干湿期的统计分布, 在计算机上进行随机抽样, 可生成干湿期随机变量值序列。这个过程也称蒙特卡罗模拟(Monte Carlo Simulation)。

已知其统计分布, 对随机变量进行直接抽样, 是常用的一种随机抽样方法。

设  $\zeta$  是某随机变量, 且其分布函数  $F(x)$  连续。

因分布函数  $F(x)$  是在  $[0, 1]$  上取值、且单调递增的连续函数, 所以, 如令随机变量  $R = F(\zeta)$ , 当  $\zeta$  在  $(-\infty, x)$  内取值时, 则  $R$  在  $[0, F(x)]$  内取值, 且

对应于 $[0,1]$ 上的一个 $r$ 值,至少有一个 $x$ 满足

$$r = F(x) = P\{\zeta < x\}$$

用 $F_1(r)$ 表示随机变量 $R$ 的分布函数,则有

$$F_1(r) = P\{R < r\} = P\{F(\zeta) < r\} = \begin{cases} 0, r \leq 0 \\ P\{\zeta < F^{-1}(r)\} = r, 0 < r \leq 1 \\ 1, r > 1 \end{cases} \quad (3)$$

故 $R$ 均匀分布在 $[0,1]$ 上。

任一随机变量 $\zeta$ 和均匀分布随机变量 $R$ 的上述关系,是通过随机数 $r$ 对随机变量 $\zeta$ 进行直接抽样的理论基础。

在统计获得干、湿期累积频率分配表之后,再应用计算机里的随机数生成函数,生成 $[0,1]$ 均匀分布随机数,进而对干(或湿)期随机变量进行如图2所示的直接抽样,即可生成有关随机变量的模拟值。

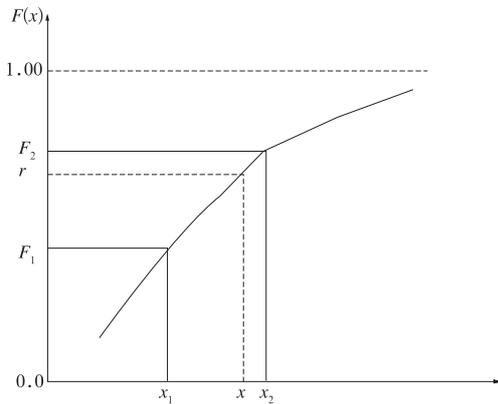


图2 基于经验分布的直接抽样  
Fig.2 Direct sampling based on empirical distribution

图2中, $F(x)$ 是干湿期的累积频率; $r$ 是 $(0,1)$ 均匀分布随机数,一般可由计算机生成,也可用专门的算法产生。当然,两者生成的都是伪随机数; $F_1, F_2$ 是紧邻 $r$ 的两个累积频率多边形节点; $x_1, x_2$ 是随机变量观测值某分组区间端点值,它们分别对应于 $F_1, F_2$ ;  $x$ 是随机变量抽样值。

因此,不难导出随机变量抽样公式为:

$$x_m = g_{j-1}^m + \Delta^m \times (r - F_{j-1}^m) / (F_j^m - F_{j-1}^m) \quad (4)$$

式(4)中, $x_m$ 为 $m$ 月随机变量(干、湿期长度)抽样值; $r$ 为 $[0,1]$ 均匀分布随机数; $F_{j-1}^m, F_j^m$ 分别为模拟参数表中最接近 $r$ 的第 $j-1$ 和第 $j$ 个累积频率值; $g_{j-1}^m$ 为模拟参数表中与第 $j-1$ 累积频率相对应

的随机变量值; $\Delta^m$ 为模拟参数表中与上述两个累积频率相对应的随机变量值差。

逐日天气序列的生成算法,如图3所示。由该算法可知,在生成湿期的同时,还相应地生成日降水量。

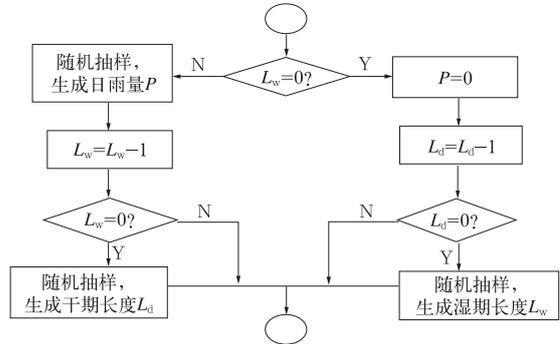


图3 基于干湿期的逐日降水生成算法  
Fig.3 Algorithm of generating daily rainfall based on dry and wet spells

### 3 DWS 模型测试<sup>①</sup>

测试所用数据为北京、太原和郑州 1951—2005 年逐日降水资料。设日降水量 $\geq 0.1$  mm 为湿日,反之则为干日。首先统计各月干湿期观测值、各月干湿期频数和累积频率分配(即干湿期的经验分布),并将其存入数据库。为建立累积频率分配表(即模型参数表),实践中,对两种资料分组方案都做了试验。因不同气候区,干、湿期统计特性各异,同一地点,不同季节,其统计特性也有明显差异,故需分站按月计算干、湿期模型参数。然后应用计算机提供的 $(0,1)$ 均匀分布随机数生成函数和式(4),按图3,依日、月、年时序,对干(或湿)期进行随机抽样,如此生成100年干湿期模拟序列。为了查明这个序列的可用性,将以月气候观测值为参考,对模拟实测值进行差异显著性检验。所涉项目包括最长干期、最长湿期、平均干期、平均湿期和平均降水日数(表1)。

另外,根据同样的历史逐日降水资料,应用 Mc 模型,统计各月干湿日转移概率,用以生成100年干湿日模拟序列<sup>②</sup>,计算上述各统计量,求其和实测值差,最后与DWS的有关统计结果一起列于表2。

由表1可见:北京、太原和郑州各月平均干期误差(即模拟值和实测值之差,下同)一般小于1 d;最

① 取自 DWS 模拟器输出,该模拟器是 DWS 的 Java 实现。

② 借助于逐日天气模拟系统 WG4E,它集成了 WGEN (Richardson C W, Wright D A, 1984)和 Shu Geng(1986)的转移概率计算模型,用 C 语言实现。

表 1 模拟与实测降水序列统计平均值差异显著性  $t$ -检验 ( $t_{0.01}=2.819, d_f=22$ )  
 Table 1 Significance test of difference between simulated and observed daily rainfall

站点	月序	最长干期/ d		平均干期/ d		最长湿期/ d		平均湿期/ d		降水日数/ d	
		实测值	模拟值	实测值	模拟值	实测值	模拟值	实测值	模拟值	实测值	模拟值
北京	1	98	94	20.7	18.6	6	4	1.5	1.4	2.0	2.4
	2	53	59	12.3	12.4	5	5	1.6	1.6	2.8	2.8
	3	46	46	9.2	9.0	4	5	1.5	1.5	3.7	3.2
	4	34	34	7.8	7.4	11	10	1.6	1.5	4.8	4.9
	5	34	34	6.2	5.8	4	4	1.5	1.5	6.0	5.8
	6	23	23	3.7	3.7	7	7	1.9	1.9	9.7	9.5
	7	14	14	2.7	2.6	14	14	2.3	2.2	13.9	14.0
	8	28	27	3.3	3.0	8	8	2.1	2.0	12.7	12.7
	9	39	39	5.6	5.3	6	6	1.7	1.6	7.4	7.5
	10	114	109	8.8	8.3	6	6	1.6	1.5	5.5	5.2
	11	93	93	15.0	13.5	6	6	1.5	1.5	3.8	3.8
	12	91	86	19.1	18.3	6	5	1.4	1.3	1.8	2.0
$t$		0.055		0.228		0.212		0.637		0.015	
太原	1	64	63	18.1	16.8	4	4	1.5	1.5	2.0	2.0
	2	45	44	12.4	12.1	5	5	1.6	1.7	2.9	3.3
	3	38	38	9.0	8.4	8	8	1.6	1.6	4.3	4.2
	4	30	29	7.0	6.8	6	6	1.6	1.5	4.9	5.2
	5	29	30	5.8	5.7	7	7	1.5	1.5	6.3	6.2
	6	17	26	3.7	3.9	8	8	1.9	1.8	9.9	9.0
	7	16	16	2.9	3.0	8	8	2.1	2.1	13.0	12.2
	8	25	25	3.3	3.2	10	10	2.0	1.8	11.9	11.5
	9	43	41	5.0	4.7	9	9	1.9	2.0	8.8	9.2
	10	143	137	9.6	8.6	7	7	1.7	1.7	5.9	6.0
	11	80	80	18.7	16.7	6	6	1.7	1.6	3.9	3.7
	12	70	72	18.0	16.8	4	5	1.6	1.5	1.8	2.2
$t$		0.006		0.237		0.111		0.573		0.044	
郑州	1	68	64	13.1	12.0	7	7	1.8	1.8	3.3	3.6
	2	35	35	8.8	8.0	6	6	2.1	1.9	4.3	4.5
	3	34	34	6.5	6.3	8	8	1.8	1.8	6.1	6.1
	4	30	27	5.9	5.6	7	7	1.8	1.8	6.9	7.3
	5	31	31	6.2	6.3	6	6	1.8	1.7	7.0	6.7
	6	22	26	5.1	4.9	9	9	1.9	2.1	7.7	8.4
	7	23	23	3.5	3.3	11	10	2.2	2.1	12.1	11.6
	8	25	25	3.9	3.8	13	13	2.0	1.9	10.6	10.7
	9	38	38	5.5	5.0	9	9	2.2	2.3	8.8	9.7
	10	59	58	8.1	7.5	9	9	2.1	2.3	7.0	7.9
	11	96	95	12.5	12.1	7	8	2.0	2.1	5.7	5.6
	12	93	82	13.8	13.2	8	8	2.0	1.9	3.7	4.0
$t$		0.129		0.291		0		0.111		0.214	

长干期误差虽普遍小于 2 d,但各站分别在 2 月、6 月和 12 月相对较大,郑州 12 月最大,达 11 d;不过,该月的平均干期误差小于 1 d。除北京 1 月最长湿期误差为 2 d 外,其他月份和太原、郑州各月湿期误差均小于 1 d,无论最长湿期或平均湿期。3 站各月平均降水日数误差均小于 1 d。 $t$ -检验显示,在  $\alpha=0.01$  的水平上,模拟值和实测值的差异是不显著的。

由表 2 可见,DWS 模拟误差普遍小于 Mc 模

型。虽然后者也能模拟出干湿期的季节变化,但绝对误差大;最长干、湿期的模拟效果,远逊于 DWS。如北京 10,11 月和太原 12 月,最长干期误差竟分别高达 47,69 d 和 41 d,郑州 11 月误差也有 23 d。月降水日数的模拟,两者效果相近,但是 Mc 的误差也相对偏大;北京 7,8 月和郑州 8 月的误差均大于 1 d。

此外,对模拟的干湿期均方差和累积频率也做了统计检验,模拟和实测值均无显著差异,限于篇

表 2 DWS 和 Mc 模拟值的绝对误差

Table 2 Absolute deviations of simulated values with DWS and Mc

站点	月序	最长干期/ d		平均干期/ d		最长湿期/ d		平均湿期/ d		降水日数/ d	
		DWS	Mc	DWS	Mc	DWS	Mc	DWS	Mc	DWS	Mc
北京	1	4	20	2.1	0.7	2	1	0.1	0.2	0.5	0.5
	2	6	16	0.1	1.8	0	0	0.0	0.0	0.0	0.1
	3	0	8	0.2	0.7	1	3	0.0	0.1	0.5	0.5
	4	0	16	0.4	0.1	1	5	0.1	0.1	0.1	0.3
	5	0	5	0.4	0.0	0	2	0.1	0.0	0.2	0.3
	6	0	7	0.0	0.2	0	1	0.1	0.1	0.2	0.7
	7	0	1	0.2	0.3	0	0	0.2	0.3	0.1	1.2
	8	1	8	0.2	0.3	0	1	0.0	0.1	0.0	1.2
	9	0	10	0.3	0.4	0	3	0.1	0.1	0.1	0.4
	10	5	47	0.5	0.9	0	1	0.1	0.1	0.4	0.5
	11	0	69	1.5	2.4	0	0	0.1	0.1	0.0	0.6
	12	5	2	0.8	6.6	1	1	0.1	0.1	0.2	0.3
	平均	1.8	17.4	0.6	1.2	0.4	1.5	0.1	0.1	0.2	0.6
太原	1	1	0	1.3	1.1	0	0	0.1	0.1	0.0	0.3
	2	1	3	0.3	0.2	0	1	0.1	0.1	0.4	0.1
	3	0	5	0.6	0.3	0	2	0.0	0.1	0.1	0.1
	4	1	10	0.2	0.7	0	1	0.1	0.1	0.3	0.2
	5	1	1	0.1	0.3	0	0	0.1	0.1	0.0	0.2
	6	9	4	0.2	0.3	0	3	0.1	0.1	0.9	0.7
	7	0	3	0.1	0.1	0	1	0.0	0.2	0.9	0.5
	8	0	4	0.1	0.1	0	1	0.2	0.1	0.4	0.7
	9	2	8	0.3	0.5	0	1	0.1	0.1	0.5	0.5
	10	6	33	0.9	0.7	0	4	0.0	0.1	0.1	0.2
	11	0	24	2.0	3.2	0	1	0.1	0.1	0.2	0.1
	12	2	41	1.2	8.2	1	2	0.1	0.1	0.4	0.2
	平均	1.9	11.3	0.6	1.3	0.1	1.4	0.1	0.1	0.3	0.3
郑州	1	4	3	1.2	0.7	0	2	0.1	0.0	0.3	0.1
	2	0	13	0.7	1.0	0	4	0.1	0.2	0.1	0.1
	3	0	0	0.2	0.5	0	1	0.0	0.0	0.0	0.4
	4	3	5	0.3	0.3	0	0	0.0	0.0	0.4	0.3
	5	0	20	0.1	0.1	0	6	0.1	0.1	0.3	0.1
	6	4	14	0.2	0.1	0	5	0.1	0.0	0.7	0.1
	7	0	3	0.2	0.1	1	0	0.1	0.2	0.6	1.0
	8	0	3	0.1	0.2	0	4	0.1	0.2	0.1	1.1
	9	0	7	0.5	0.6	0	2	0.1	0.2	0.9	0.8
	10	1	15	0.5	0.9	0	1	0.2	0.1	0.9	0.2
	11	1	23	0.4	2.4	1	1	0.1	0.4	0.1	0.7
	12	11	19	0.6	2.3	0	1	0.1	0.0	0.3	0.2
	平均	2	10.4	0.4	0.8	0.2	2.3	0.1	0.1	0.4	0.4

幅未列表内。

DWS 模拟器实际还模拟生成了湿期内的逐日降水量,并且同样效果良好,因本文仅讨论干湿期直接相关量,故也未列表内。

#### 4 讨 论

测试结果显示,所建干湿期模型是可用的。

DWS 模型性能比 Mc 模型相对较好。主要原

因在于,前者对于干(或湿)期变量直接建模,而后者定义的随机变量是干湿日状态,它不能“识别”干湿期。为了生成干湿日状态的时间序列,Mc 总是以前 1 日的干湿状态为条件,按转移概率进行随机抽样。且每生成 1 个干(或湿)日,总不能预知下 1 日状态;下 1 日状态将由下 1 日的随机抽样来决定。至于干湿期,则是在已生成干湿日序列之后被测定的。如测定出一个长度为  $N$  天的干期,则它必是  $N$  次随机抽样的结果。

在 DWS 模型中, 随机变量干(或湿)期, 由经验分布函数给出其完整的描述, 干湿期序列的生成, 是依据分布函数对干(或湿)期变量作直接的随机抽样。每生成 1 个干(或湿)期, 可预知下一步一定是生成 1 个湿(或干)期。显然, 若生成一个长度也是  $N$  天的干(或湿)期, 它只是一次抽样的结果。

另外, Mc 模型建立在一个假设条件之下, 即每次干湿日状态的转移, 仅仅依赖于前一次转移后的状态, 而与更以前的情况无关。这和天气变化的实际并不吻合。DWS 模型基于经验分布函数, 它不使用任何理论假设, 唯一地依赖于天气资料样本。

虽然基于经验分布函数的 DWS 模型有较好的模拟效果, 但付出的代价是用了较多的参数。在建立经验分布时, 倘数据样本被分为  $N$  组, 则将产生  $2N+1$  个模型参数, 从数学角度看, 缺少数学美感。反之, Mc 模型仅用了两个模型参数。

鉴于 DWS 模型的可用性, 它已作为核心部件被耦合到多要素随机天气模型中。有关细节, 将另文讨论。

**致 谢:** 本文所用气象资料承蒙国家气象信息中心提供, 作者谨表诚挚谢意。

### 参 考 文 献

- [1] Richardson C W, Wright D A. WGEN: A Model for Generating Daily Weather Variables. USDA, ARS-8 Washington D C, 1984.
- [2] Johnson G L, Hanson C L, Hardegree S P, et al. Stochastic weather simulation—Overview and analysis of two commonly used models. *J App Meteorol*, 1996, 35: 1878-1896.
- [3] Larsen G A, Pense R B. Stochastic simulation of daily climatic data for agronomic models. *Agronomy Journal*, 1982, 74: 510-514.
- [4] 林而达, 张厚宣, 王京华, 等. 全球气候变化对中国农业影响的模拟. 北京: 中国农业科技出版社, 1997: 23-53.
- [5] Semenov M A, Brooks R J, Barrow E M, et al. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research*, 1998, 10: 95-107.
- [6] 王世耆. 随机天气模拟原理与模拟系统 WG4E. 计算机与农业(增刊). 作物管理软件研究专集, 1997: 67-82.
- [7] Shu Geng. A Simple Method for Generating Daily Rainfall Data. *Agricultural and Forest Meteorology*, 1986: 363-376.
- [8] Richardson C W. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 1981, 17(1): 182-190.
- [9] Jorgenson D L. Persistency of rain and no-rain periods during the winter at San Francisco. *Mon Wea Rev*, 1949, 77(9): 303.
- [10] Steinhauser H. Trocken- und Niederschlagsperioden und ihre Theoretische Behandlung. *Arch Met Geoph Biokl*, 1959, 10(1): 38-58.
- [11] Gabriel K R, Neumann J. On distribution of weather cycles by length. *Quart J R Met Soc*, 1957, 83: 357-375.
- [12] Longley R W. The length of dry and wet periods. *Quart J R Met Soc*, 1953, 79: 342-520.
- [13] 么枕生. 气候统计. 北京: 科学出版社, 1963: 170-172.
- [14] William H P, Brian P F, Saul A T, et al. Numerical Recipes in C: The Art of Scientific Computing. Cambridge: University Press, 1988: 517-518.
- [15] 诸叶平, 王世耆. 随机天气模型及其 JAVA 实现. 电子学报, 2007, 35(12): 25-29.
- [16] 复旦大学数学系. 统计数学. 上海: 上海科学技术出版社, 1960: 67-69.

## Stochastic Simulation for Dry and Wet Spell

Wang Shiqi Zhu Yeping Li Shijuan

(*Institute of Agricultural Information, CAAS, Beijing 100081*)

### Abstract

Rainfall models are the most important component in stochastic weather generator. Two-state, first-order Markov chain model is generally applied to simulate rainfall occurrence. The monthly statistics of time series of dry and wet days simulated by the model shows it may work well, but it is not satisfying when focusing on the persistent drought or prolonged wet in the series, although the difference between the simulated monthly mean of rainy days and the actually observed one are not marked.

A stochastic model of dry and wet spells (DWS) is described, in which defined stochastic variables are the length of dry or wet spells, numbering in days, other than dry and wet day state. It is obvious that the variable itself has expressed the persistency of rainy or drought weather. Data modeling method is applied too. The related techniques include designing an algorithm for obtaining observed data of dry and wet spells from history records of daily rainfall; constructing empirical distribution function of the length of dry and wet spells monthly, and creating the parameter tables mapping the accumulated frequency distribution monthly; deriving a stochastic sampling formula for generating a dry or wet spell based on direct sampling principle and an algorithm of daily weather(dry or wet) on computer based on Monte Carlo simulation technique with previous sampling formula and parameter tables. Dry and wet spell simulation has been implemented using Java language. Users can select some run time parameters, for example, the name of observed location, the thread value for rainy day, and so on.

Model validation test are done using history data from three locations, Beijing, Taiyuan and Zhengzhou. 100 years of rainfall data are generated for each location with the help of DWS simulator respectively. Its statistic items monthly includes: maximum of spell, mean of spell, variance of spell and mean number of rainy days. The mean absolute deviation of simulated value from observed one for all statistical items are about 1.8—2.0, 0.1—0.4, 0.4—0.6, 0.08—0.09 and 0.2—0.4, respectively.

The *t*-tests are done in order to detect significant differences between observed and simulated value for maximum, mean and variance. No significant differences are found at  $\alpha=0.01$ . For comparison between dry and wet spell model and two-state, first-order Markov chain model, the same statistics are obtained by running Markov chain model. Results indicate that the accuracy of dry and wet spell model is higher than two-state, first-order Markov chain for all statistical items, especially for maximum dry spells.

Although dry and wet spell model is available and better than two-state, first-order Markov chain, its weakness is that the parameters in dry and wet spell model are more than those in Markov chain model, lacking in aesthetic feeling of mathematics.

**Key words:** dry and wet spell; stochastic modeling; weather simulator