

随机天气模型及其 JAVA 实现

诸叶平,王世耆

(中国农业科学院农业信息研究所,北京 100081)

摘 要: 应用逐日天气随机模型,设计和开发了基于 JAVA 的逐日降水、温度和太阳辐射多要素随机模拟器 DWSS,论述其功能实现和主要特性,并列举了应用例子。DWSS 能够跨平台运行,可以通过浏览器实现远程访问,不仅生成逐日天气随机序列,还向用户提供更多的选择。

关键词: 天气;随机模型;Java;Web

中图分类号: TP399 **文献标识码:** A **文章编号:** 0372-2112 (2007) 12-2267-05

Stochastic Modeling of Daily Weather and Its Implementation in Java

ZHU Ye-ping, WANG Shi-qi

(Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: A stochastic simulator prototype DWSS on precipitation, temperature and solar radiation in JAVA was developed. This paper describes the designing considerations, functions, and features of it. DWSS can be used on desktop or network environment, and a running example with the help of some figures and table on desktop is presented. This simulator can provide more selections to users than other popular weather generators.

Key words: weather; stochastic modeling; Java; Web

1 引言

天气要素的长序列值,是很多研究和风险评估模型中不可缺少的输入数据。气象台站的逐日天气序列实测值,只是逐日天气随机变化的一个样本,经常不能满足模型输入的需要。有些地方则根本没有观测值可用。实测资料应用的时空局限性,导致模拟数据的兴起。

依据概率统计原理,模拟天气要素长序列值的方法,自上世纪 70 年代初,即已在水文学研究中获得应用^[1,2]。稍后,Crank, K. N. (1977) 提出模拟逐日天气随机值的方法^[3]。Douglas C. Bond(1979)在此基础上,用 80 年历史观测数据,估计 5 - 8 月生长季每日天气变量的概率统计参数,并用这些参数和 Monte Carlo 模拟技术在美国农业部华盛顿计算中心的 IBM 370/168 机上,生成 100 年长度的逐日降水、最高和最低温度模拟值,为当时的作物产量概率统计预报提供输入数据^[4]。这类天气模拟程序又称之为天气生成器或天气发生器(Weather Generator)。上世纪末,以气候变化为核心的全球环境问题的研究导致长序列逐日天气数据的需求增长,这些随机生成或模拟技术遂引起人们的关注。

Richardson, C. W. (1984) 发表了一个实用性较强的模型——逐日天气生成器 WGEN。他用遍布全美 48 个州的 139 个台站资料,统计出 48 个雨日、雨量参数和 12 个温度、辐射参数,且制成参数列表和等值线图。WGEN 可生成降水、最高气温、最低气温和太阳辐射变量的逐日序列^[5]。其后几个基于 WGEN 的天气生成器,如 USCLIMATE、CLIMGEN^[6]和 CWG^[7]等相继发表。

估算 WGEN 模型的雨日、雨量参数,必须有长序列逐日气象记录。Shu Geng 研究了用常规月气候统计值,估算参数的方法,简称短序列方法^[8]。作者曾用北京 28 年资料对此作过测试^[9]并开发了兼具两种参数估算方法的天气模拟系统 WG4E。

Racsko 等(1991)发展了另一类天气生成模型 LARS-WG,它对降水变量的处理与 WGEN 不同,所模拟的不是干日和雨日,而是干期和雨期^[10]。Simenov, M. A. 和 Brooks, R. J. 等(1998)对 WGEN 和 LARS-WG 用不同气候区的美国、欧洲和亚洲的 18 个站点资料,进行了测试和比较^[11]。廖要明等(2004)用一阶马尔可夫链和 Gamma 分布开发了降水天气发生器 NCC/RCGW^[12]。

已发表的随机天气生成器,虽因研究目的不同,采

用的数学模型各异,但提供给用户的功能,都大致相似:可运行在 Windows 平台上,单机操作以获取任意长度的逐日天气数据.一个地点具有一组模拟参数,因此,生成的逐日天气数据,反映某地以最大长度的历史逐日天气资料所估计的平均气候特征,如长年月平均值或方差等.

然而,实践表明,气候是变化的.在不同历史时段的逐日天气数据中,寓藏着这种气候变化信息.研究未来气候变化的影响,是现代社会经济和科学研究的重要课题.在现实中,人们不仅需要符合长年气候特征的长序列逐日天气数据,更需要能反映长序列中某一时段气候特征的长序列逐日天气数据,而且这些数据必须能为用户共享,这是研发 DWSS 的缘起. DWSS 允许用户按需随机生成逐日天气数据,并具有跨平台应用和资源共享的特性.

2 逐日天气随机模型

构造逐日天气随机模拟器的重要基础是逐日天气变量的概率统计模型或随机模型. DWSS 所用的数学模型如下:

2.1 均匀马尔可夫链——干、雨日转移模型

逐日降水可视为一个序列试验

$$E_0^0 E_0^1 E_1^2 \dots E_i^s E_j^{s+1} \dots E_j^{s+n} \dots$$

其中 E_i^s 表示降水事件,下标 i 表示降水量级别(设有 $k+1$ 级),上标 s 表示试验序号.如果在第 s 次试验出现事件 E_i^s 的条件下,第 $s+1$ 次试验出现事件 E_j^{s+1} 的条件概率与试验序号无关,则称此试验序列形成一个均匀马尔可夫链.而所说的条件概率称之为转移概率,记为 p_{ij} . 又令 $p_{ij}(n)$ 表示在第 s 次试验出现事件 E_i^s 的条件下,经 n 次试验转为 E_j^{s+n} 的概率,则发生该转移的全部概率构成一个转移矩阵

$$M_n = \begin{bmatrix} p_{00}(n) & p_{01}(n) & \dots & p_{0k}(n) \\ p_{10}(n) & p_{11}(n) & \dots & p_{1k}(n) \\ \dots & \dots & \dots & \dots \\ p_{k0}(n) & p_{k1}(n) & \dots & p_{kk}(n) \end{bmatrix} \quad (1)$$

根据全概率公式,

$$M_n = M_1^n$$

这里, n 称之为马尔可夫链的阶.

有些研究者认为,用马尔可夫链模拟逐日降水,阶应不小于 2;但许多实践证明,若仅用于干、湿两状态的模拟,可以只用 1 阶.

在 DWSS 中,雨日的发生和雨量大小是分别建立模型的降水量按是否大于 0.1 毫米,被分成雨日和干日二个级别.和唯一使用马尔可夫链模拟逐日降水不同,这里只用一阶马尔可夫链单独模拟雨日的发生.因此,

式(1)被简化为

$$M_1 = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \quad (2)$$

式中 p_{ij} 的下标 0 表示干日,1 表示雨日.

因为在转移矩阵中,每行之和等于 1,故对于式(2),只要统计出右列两项转移概率值,则左列另两项概率值很容易算得.

2.2 两参数 Gamma 分布——日雨量模型

Gamma 概率密度函数的一般形式为:

$$f(x) = \frac{(x-c)^{-1} e^{-(x-c)/\lambda}}{\Gamma(\lambda)}$$

式中 x 为随机变量, c 为 x 的下限值, λ 和 $\Gamma(\lambda)$ 为分布参数, $\Gamma(\lambda)$ 为 Gamma 函数.若 x 表示日雨量,可设 $c=0$. 此时两参数的 Gamma 分布有下列形式的概率密度函数:

$$f(x) = \frac{x^{-1} e^{-x/\lambda}}{\Gamma(\lambda)} \quad (3)$$

用 Gamma 分布模拟日雨量,须估计两个参数 λ 和 $\Gamma(\lambda)$. 其估算公式为

$$\lambda = (0.5000876 + 0.16488552y - 0.0544274y^2) / y \quad (4)$$

$$\Gamma(\lambda) = \bar{x} / \lambda \quad (5)$$

式中 $y = \ln(\bar{x}/g)$. \bar{x} 是日雨量 x 的算术平均数, g 是日雨量的几何平均数 (G. A. Larsen and R. B. Pense, 1982) ^[13].

2.3 一阶线性自回归——日辐射和温度模型

太阳辐射、最高和最低气温的逐日变化,被视为连续多维(多变量)随机过程.变量间的序列相关可以用一阶自回归模型来描述 (C. W. Richardson, 1981) ^[14]. 其基本表达式为

$$p_{,i}(j) = A_{p,i-1}(j) + B_{p,i}(j) \quad (6)$$

式中 $p_{,i}(j)$ 和 $p_{,i-1}(j)$ 都是 (3×1) 矩阵,其元素是日最高气温、日最低气温和日太阳辐射的残差; $p_{,i}(j)$ 是由正态分布的独立随机分量构成的 (3×1) 矩阵,每个分量具有平均数为 0,方差为 1 的正态分布. A 和 B 是 (3×3) 矩阵,其元素由变量的后延自相关和互相关系数构成. $j=1,2,3$ 分别代表日最高气温、日最低气温和日太阳辐射;下标 p 为年序, i 为日序.

残差变量值由下列公式生成:

$$p_{,i}(j) = \frac{X_{p,i}(j) - \bar{X}_i^0(j)}{\sigma_i^0(j)} \quad \text{当 } R_{p,i} < 0.1\text{mm, 即干日} \quad (7)$$

或

$$p_{,i}(j) = \frac{X_{p,i}(j) - \bar{X}_i^1(j)}{\sigma_i^1(j)} \quad \text{当 } R_{p,i} \geq 0.1\text{mm, 即雨日} \quad (8)$$

式中, $X_{p,i}(j) - \bar{X}_i^0(j) - \bar{X}_i^1(j) - \sigma_i^0(j) - \sigma_i^1(j)$, $R_{p,i}$ 是年、日编号分别为 p, i 时的日降水量, $X_{p,i}(j)$: 对应于

$p_{,i}(j)$ 的原变量值, $\overline{X_i^0(j)}$:原变量在干日的平均数, $\overline{X_i^1(j)}$:原变量在雨日的平均数, $\sigma_i^0(j)$:原变量在干日的标准差, $\sigma_i^1(j)$:原变量在雨日的标准差。

式(7)和(8)的意义在于剔除有关变量的季节变化和降水因素的影响,突显其随机变化成分。又因温度和辐射的年变化属于复杂的周期现象,在理论上可用周期函数来描述,故式中各变量平均数与标准差的年变化实际是时间(日)的周期函数,可用谐波分析。这使模型的表达更为简洁。

在模型中,辐射、最高、最低气温和降水 4 个变量,有独立变量和非独立变量之分。降水属于前者,最高、最低气温和辐射属于后者。在建模和估计模型参数时,非独立变量的处理,总以独立变量的取值为前提。在逐日天气序列随机生成时,必须先模拟降水,后模拟温度和辐射,即必须先独立地生成降水变量值,再以其为背景,条件地生成最高、最低气温和太阳辐射值。

3 DWSS 的设计、功能和特性

DWSS 的主要任务是,根据用户需求,生成具有特定气候特征的逐日天气变量值。这意味着,系统应具有选择所需历史数据,实时(或在线)统计随机模拟参数和实时(或在线)生成逐日天气数据的功能。为此, DWSS 的设计,主要关注以下几点:

模拟器所涉历史天气数据,由关系数据库管理。以方便从中提取感兴趣的、具有某类气候特征的历史数据。

历史天气数据的统计分析和模拟参数的计算结果,均对用户可视,使模拟器不局限于仅提供随机生成的逐日天气序列。

模拟器实现要利于系统扩展、移植和跨平台运行。

模拟器网络应用的客户端平台采用浏览器。WEB 应用程序采取 SERVLET 方案,以便使 DWSS 具有更高的运行效率。

模拟器可同时适应单机和网络应用环境。

基于上述构想,更由于 Java 提供了一个概念清晰、结构紧凑的分布式计算模型和互操作的计算方法^[15],故选择 Java 实现 DWSS 及其网络应用。其中 DWSS 含有 49 个 JAVA 类,分别实现连接数据库、气候统计分析、模拟参数计算、随机生成逐日天气等。

已实现的 DWSS 可提供下列功能:

(1) 逐日天气统计分析 可对任意多个站点的不同序列长度的逐日天气资料,进行谐波(富里哀级数)分析和一阶马尔可夫链统计。

(2) 随机模拟参数生成 可生成多站点的标准模拟参数。也可据用户输入,生成特定的模拟参数。

(3) 逐日天气随机模拟 可据用户选择,或用标准模型参数,或用定制模型参数,模拟任意长度的逐日随机天气资料。

以下是 DWSS 的一些重要特性:

(1) 可重用性好,易于移植和嵌入其他应用系统。

(2) 可运行在 J2EE WEB Container 环境下,用户端可通过浏览器输入数据、启动运行模拟器和获取所需结果。

(3) 可运行在多个不同操作平台上,如 WINDOWS、MAC、SOLARIAS、Linux、及各种 FreeBSD 等。

(4) 不同用户可共享逐日天气数据库的数据资源。对数据库的访问使用 JDBC 2.0 标准接口,兼容几乎所有 RDBMS 平台。

(5) WEB 用户界面采用 SERVLET 标准,可运行于所有 J2EE1.3 兼容的应用服务器运行环境。

4 DWSS 网络应用框架

天气模拟器的最终目的是:借助它,获取逐日天气变量值,以便为一些应用程序,提供必要的输入数据。DWSS 网络应用采取的结构框架如图 1。

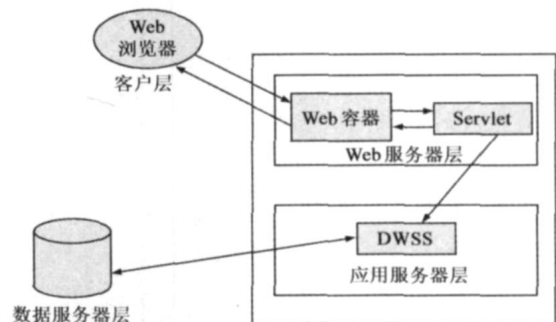


图 1 DWSS 应用框架

在这个框架里,浏览器是用户输入需求信息、向远程服务器请求运行随机模拟器,并获取模拟结果数据的工作平台。

Web 服务器存放将被浏览的页面,处理用户请求和调用 Servlet。

Servlet 的任务包括:处理用户信息;访问数据库;生成提示信息,通过 Web 服务器,反馈给浏览器;将用户信息传递给 DWSS,并运行后者;把 DWSS 生成的逐日天气随机序列值及其他有关信息,存入数据库,根据模拟结果,构造 Web 页面,以备用户浏览;生成含提示下载信息的 Web 页面,备用户点击下载,等等。

DWSS 是应用框架的内核。它统计逐日天气数据,计算模型参数,生成天气随机序列,并将结果写入数据库。

数据库是所有有用信息存放地。包括历史天气资料、逐日天气统计值、模型参数和天气变量的随机模拟

值序列。

基于上述应用模式,用户无需安装天气生成器,即可从网上获取任意长度的天气随机值序列。不存在本机运行的设备配置问题,客户端虽无历史天气数据源,也可共享远程数据库里的数据资源,按需统计模型参数,并用这些参数,生成天气随机序列,这有利于模型研究和模拟器更新。

因为不同用户通过浏览器按需输入数据,获取模拟结果时,模拟器每次运行的全部数据,包括模型参数值,也集中存储在数据库里,模拟器研究人员可以方便地存取这些数据,以便分析模型应用效果、改进模型和随时更新模拟器。

5 应用举例

以下简述 DWSS 的应用。假设需要以北京 1960-1980 这段时期的天气气候为背景,构造 100 年的逐日最高、最低气温,太阳总辐射和降水,作为某个应用程序(如作物模型)的输入数据。

表 1 随机模拟结果统计

月序		1	2	3	4	5	6	7	8	9	10	11	12
雨日	模拟值	2.1	2.9	4.4	4.5	5.3	9.4	13.7	12.1	7.7	5.8	3.8	1.5
	实测值	1.9	2.8	3.6	4.8	5.4	9.3	13.8	12.4	7.5	6.2	4.0	1.5
雨量 (mm)	模拟值	4.0	7.3	8.5	24.4	22.7	51.4	92.0	195.2	254.8	18.5	6.1	2.3
	实测值	2.8	6.4	7.9	22.8	23.1	65.6	90.2	185.9	251.1	22.2	5.1	2.4
平均最高 温度(度)	模拟值	1.8	4.6	11.3	20.3	26.5	30.5	31.1	29.5	25.6	18.5	9.7	3.1
	实测值	1.5	4.1	11.2	19.8	26.4	30.4	31.0	29.5	25.7	18.9	9.9	3.1
平均最低 温度(度)	模拟值	-9.3	-6.6	-0.5	7.3	13.1	18.2	21.8	20.2	13.9	7.0	-0.7	-7.2
	实测值	-9.7	-7.1	-0.7	7.0	13.0	18.1	21.7	20.3	14.0	7.1	-0.6	-7.1
平均太阳 总辐射 (MJ/m ² .d.)	模拟值	9.0	11.4	15.0	18.6	21.5	22.2	20.4	18.4	15.3	11.7	9.0	8.3
	实测值	9.0	11.8	15.8	18.9	21.9	21.8	17.3	16.3	12.6	9.1	7.8	

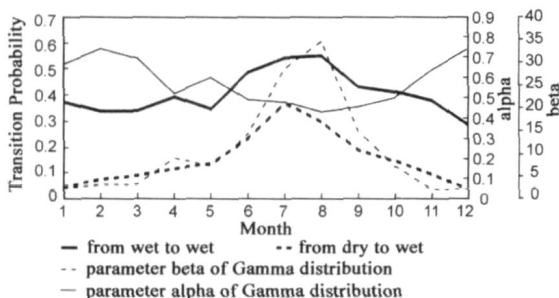


图 2 降水模拟参数

首先,用户通过提供用户名和口令等连接到相应的数据库系统,然后,按提示和应用需求,选择输入,并定义一个输出文件名。点击“启动随机模拟器”按钮,DWSS 即自动从相应数据库中读取 1960 - 1980 年历史天气数据、统计模拟参数、随机生成天气变量值,并把统计和模拟结果写入数据库,同时,弹出一个窗口,显示逐日最高、最低气温,太阳总辐射和降水的随机生成

值,这些数据又被写入用户定义的输出文件中。

作业完毕,DWSS 在数据库中保存有与该此作业有关的全部信息。

图 2 是用北京 1958-1985 年资料计算的 4 个降水模拟参数。表 1 是逐日天气生成值的统计结果。

6 讨论

逐日天气的随机模拟,依赖于有关变量的统计学知识,尤其是有关变量的分布函数。从应用角度说,逐日天气的随机模拟,可应用理论分布函数,也可使用随机变量频数分配表。就模拟器构造而言,前者是一个计算过程,后者是一个搜索过程。如用频数分配表,且将历史资料分为 N 组,则将产生 $2N + 1$ 个参数,显然远多于理论分布函数中的参数个数。当然,使用的参数愈少,模型愈能反映事物的本质。

滤掉季节变化的多天气变量随机模型,除应用(6),还可采用其他的概率数据关联算法^[16],但可预见,它们都将占用更多的计算机资源。

研究全球气候变化对一些产业的影响,一般使用长序列本地天气资料,结合使用全球气候模型(GCMs)的输出,以构造未来气候情景。但 GCMs 的输出,并非预报值。它们有某种不确定性。依据本地历史观测资料,可用 DWSS 构造某种“未来”气候情景。换言之,如果有观测资料可用,一个地点,用不同时期具有不同特征的天气观测值,可以估计出不同的“气候”统计参数。如将这些不同的统计参数输入 DWSS,则可获得不同“气候”背景下的随机天气变量值。

DWSS 的设计,是基于一种发展中的数据共享和软件共享模式。历史逐日天气数据被存放在某个数据库里,通过授权,终端用户可通过浏览器远程访问某个服务器里的随机模拟器,获取所需的信息。当然,历史天气数据结构,应由数据拥有者决定,它可能不同于我们所拟定的结构。但是,这不妨碍 DWSS 架构的可用性。

参考文献:

- [1] Khanal N N, R L Hamrick. A stochastic model for daily rainfall data synthesis [A]. 1971 Proceedings of the Symposium on Statistical Hydrology [M]. USA: USDA Miscellaneous Publication 1275, 1974. 97 - 210.
- [2] Nicks A D. Stochastic generation of the occurrence, pattern, and location of maximum amount of daily rainfall [A]. 1971 Proceedings of the Symposium on Statistical Hydrology [M]. USA: USDA Miscellaneous Publication 1275, 1974. 154 - 171.
- [3] Crank K N. Simulating Daily Weather Variables [M]. USA: Washington, D. C., USDA, Statistical Reporting Service, 1977.

- [4] Bond D C. Generating Daily Weather Values by Computer Simulation Techniques for Crop Yield Forecasting Models [M]. USA:USDA-ARS (formerly USDA-ESCS), statis Res Div, 1979. 38.
- [5] Richardson C W, Wright D A. WGEN: A model for Generating Daily Weather Variables [M]. Washington, DC: USDA, ARS-8, 1984.
- [6] Johnson G L, Hanson C L, Hardegree S P, Billard E B. Stochastic weather simulation: overview and analysis of two commonly used models [J]. J App Meteorol, 1996, (35): 1878 - 1896.
- [7] 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 [M]. 北京: 中国农业科技出版社, 1997. 23 - 53. Lin Eda, Zhang Houxuan, Wang Jinghua. Simulating the Effect of Global Climate Change on China Agriculture [M]. Beijing: Chinese Agricultural Science and Technology Press, 1997. 23 - 53. (in Chinese)
- [8] Shu Geng. Simple method for generating daily rainfall data [J]. Agricultural and Forest Meteorology, 1986, 36(4): 363 - 376.
- [9] 王世耆. 随机天气模拟原理与模拟系统 WG4E [J]. 计算机与农业, 1997, (增刊): 67 - 82. Wang Shiqi, Stochastic weather modeling and simulator WG4E [J]. Computer and Agriculture, 1997, (supplement): 67 - 82. (in Chinese)
- [10] Racsko P, Szeidl L, Semenov M. A serial approach to local stochastic weather models [J]. Ecol Model, 1991, 57(1 - 2): 27 - 41.
- [11] Semenov M A, Brooks R J, Barrow, E M, Richardson C W. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates [J]. Climate Research, 1998, 10(8): 95 - 107.
- [12] 廖要明, 等. 中国天气发生器的降水模拟 [J]. 地理学报, 2004, 59(5): 689 - 698. Liao Yaomi, et al, Precipitation simulation in China with a weather generator [J]. Journal of Geographical Sciences, 2004, 59(5): 689 - 698. (in Chinese)
- [13] Larsen G A, Pense R B. Stochastic simulation of daily climatic data for agronomic models [J]. Agronomy Journal, 1982, 74: 510 - 514.
- [14] Richardson C W. Stochastic simulation of daily precipitation of temperature and solar radiation [J]. Water Resources research, 1981, 17(1): 182 - 190.
- [15] 朱斌, 刘子欣, 朱海云. JAVA 与 CORBA 结合实现分布式异构系统中间件 [J]. 电子学报, 2003, 31(9): 1313 - 1316. Zhu bin, et al, Integrate JAVA with CORBA to implement the middleware of the distributed systems of heterogeneity [J]. Acta Electronica Sinica, 2003, 31(9): 1313 - 1316. (in Chinese)
- [16] 潘泉, 等. 广义概率数据关联算法 [J]. 电子学报, 2005, 33(3): 467 - 472. Pan quan, et al. Generalized probability data association algorithm [J]. Acta Electronica Sinica, 2005, 33(3): 467 - 472. (in Chinese)

作者简介:



诸叶平 女, 1959 年生于北京, 研究员, 博导, 农业部中青年有突出贡献专家. 1982 年毕业于天津大学计算机系, 一直从事计算机农业应用研究, 1993 年获政府特殊津贴, 曾主持国家科技攻关、国家“863”计划、农业部重点、北京市自然科学基金等课题 30 余项, 获部级以上科技成果奖 13 项, 在国内外学术刊物上发表论文 50 余篇, 主要研究方向: 信息技术应用、作物模拟模型和

智能技术应用. E-mail: zhuyep@mail.caas.net.cn



王世耆 男, 1938 年出生于江苏, 研究员, 1962 年毕业于南京大学气象系, 1991 年获政府特殊津贴, 发表论文 30 余篇. 研究方向: 计算机农业应用研究和软件开发.