

基于信息扩散理论的天津旱涝灾害危险性评估^{*}

杜晓燕, 黄岁樑, 赵庆香

(南开大学 环境科学与工程学院, 天津市城市生态环境修复与污染防治重点实验室,
环境污染过程与基准教育部重点实验室, 天津 300071)

摘要: 采用基于信息扩散理论的评估模型、应用天津近22年的旱涝指数为样本, 对天津旱涝危险性进行评估, 给出了天津旱涝指数概率密度曲线及不同程度旱涝危害发生的可能性。结果表明, 重旱、重涝几乎不可能发生; 大旱发生的可能性也很小——约700多年一遇; 偏旱、大涝、偏涝这3种强度的灾害则比较频发, 年发生概率分别为0.231 3、0.159 1、0.073 4。评估结果与天津近22年的实际情况基本吻合, 说明应用此模型, 对信息量不足的自然灾害进行危险性评估是可行的。

关键词: 信息扩散; 降水距平百分率; 旱涝灾害; 危险性评估; 天津

中图分类号: O159; P426.616 **文献标识码:** A **文章编号:** 1000-811X(2009)01-0022-04

0 引言

天津是中国北方重要的沿海开放城市, 有着重要的政治经济地位, 但又处在华北平原灾害频发地带, 水旱灾害即是其中影响天津经济发展的自然灾害。自20世纪80年代起, 全球开始增温, 引起某些区域气候特征改变。随着增温进程的延续, 华北地区干旱现象趋于严重^[1]。1980~1982年天津市连续干旱受灾面积达36万hm², 是解放以来最严重的一次干早期^[2]。同时, 天津东临渤海, 地势低平, 地下水位高, 有众多坑塘洼淀、地上河、古河道等每逢暴雨易使地表径流不畅导致涝灾; 而且天津地处海河尾闾, 有“九河下梢”之称, 客水泄洪也很可能诱发涝灾。因此天津面临着旱灾与涝灾的共同威胁。苏同卫等^[3]通过分析指出, 天津及其周围地区早年、涝年基本呈正态分布, 而且旱涝变化呈现10.1、19.8、35.7年的周期性, 与太阳活动紧密相关。

旱涝灾害是两种典型的自然灾害, 在进行危险性分析或风险评估时往往遇到历史灾情资料不足的情况。旱涝危险性评估的主要目的是给出不同强度危害发生的可能性, 即其概率密度函数。

在样本信息不足的情况下, 采用基于大数定理的传统概率论方法所得的结果必然不精确。信息扩散原理是一种处理不完备信息的模糊方法, 认为一定存在着一个适当的扩散函数, 可以将传统的观测样本点集值化, 以弥补资料不足带来的缺陷, 达到提高精度的目的^[4]。所以, 本文利用信息扩散理论的评估模型对天津不同程度旱涝危害进行评估, 从而给出不同程度旱涝危害发生的可能性, 力求得到较为反映实际情况的结果。

1 信息扩散理论评估模型

信息扩散理论由黄崇福教授在《模糊信息优化处理技术及其应用》^[4]一书中系统提出。其处理对象是小样本提供的模糊不完备信息(仅仅依靠他们, 不可能清楚地得出统计规律), 依据是关于信息扩散原理的一个断言: 假设给定了一个知识样本, 用它可以估计一个关系, 直接使用该样本得出来的结果称为非扩散估计; 当且仅当该样本不完备时, 一定存在一个适当的扩散函数和一个相应的算法, 使得扩散估计比非扩散估计更靠近真实估计^[5]。信息扩散是一种尽可能挖掘可能多的信息, 以提高系统识别精度的技术, 因此它与模糊信息矩阵、信息分

^{*} 收稿日期: 2008-06-04

基金项目: 水利部科技创新项目(XDS2007-05); 天津市科技发展计划项目(05YFSZSF02100); 国家自然科学基金项目(50479034)

作者简介: 杜晓燕(1982-), 女, 山东泰安人, 硕士, 主要从事自然灾害风险评估方面研究. E-mail: dodoyan@mail.nankai.edu.cn

通讯作者: 黄岁樑(1964-), 男, 湖南南县人, 博士(后), 教授, 主要从事水体环境学与水环境数值模拟研究。

E-mail: slhuang@nankai.edu.cn

配方法和内集——外集模型并称为模糊信息优化处理技术。近年来, 模糊信息扩散技术不仅用以完成了重大科技公关项目, 例如王新洲主持完成的国家自然科学基金项目“信息扩散原理在测量数据处理中的应用研究”^[5], 而且在自然灾害风险评价领域得到了越来越广泛地应用^[6-9]。

基于信息扩散理论的评估模型具体如下:

设灾害指数论域为

$$U = \{u_1, u_2, \dots, u_m\}, \quad (1)$$

灾害指数样本集合为

$$X = \{x_1, x_2, \dots, x_n\}, \quad (2)$$

一个单值观测样本点 x 依式(3)将所携带的信息扩散给 U 中的所有点

$$f_i(u_j) = \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(x_i - u_j)^2}{2h^2}\right], \quad (3)$$

式中: h 为扩散系数, 由样本集合中的最大值 b 、最小值 a 和样本点个数 n 确定。

$$h = \begin{cases} 0.8146(b-a) & n=5 \\ 0.5690(b-a) & n=6 \\ 0.4560(b-a) & n=7 \\ 0.3860(b-a) & n=8; \\ 0.3362(b-a) & n=9 \\ 0.2986(b-a) & n=10 \\ 2.6581(b-a)/(n-1) & n \geq 11 \end{cases} \quad (4)$$

取

$$A = \max_{1 \leq j \leq m} \{f(u_j)\}, \quad (5)$$

则

$$\mu_x(u_j) = \frac{f(u_j)}{A}, \quad (6)$$

就将单值样本点 x 变成了一个以 $\mu_x(u_j)$ 为隶属函数的模糊子集。

在进行风险评估时, 为了使每一个集值样本点的地位相同, 需对式(5)做适当的调整, 所得的模糊子集也不再是式(6)中最大隶属度为 1 的正规化模糊子集。设对第 i 个样本点 x_i 按式(3)进行信息扩散, 令

$$C_i = \sum_{j=1}^m f_i(u_j), \quad (7)$$

相应的模糊子集的隶属函数为

$$\mu_{xi}(u_j) = \frac{f_i(u_j)}{C_i}, \quad (8)$$

称 $\mu_{xi}(u_j)$ 为样本点 x_i 归一化信息分布。对 $\mu_{xi}(u_j)$ 进行处理, 便可得到一种效果较好的风险评估结果, 令

$$q(u_j) = \sum_{i=1}^n \mu_{xi}(u_j), \quad (9)$$

其物理意义是由 $\{x_1, x_2, \dots, x_n\}$ 经信息扩散推断出, 如果灾害观测值只能取 u_1, u_2, \dots, u_n 中的一个, 在将 x_i 均看作是样本点代表时(包括部分的代表在随机实验中没有出现的样本点), 观测值为 u_j 的样本点个数为 $q(u_j)$ 。显然 $q(u_j)$ 通常不是一个正整数, 但一定是一个不小于零的数。再令

$$Q = \sum_{j=1}^m q(u_j), \quad (10)$$

Q 事实上就是各 u_j 点上样本点数的总和, 从理论上讲, 应有 $Q = n$, 但由于计算过程中四舍五入的误差, 它们之间通常略有差别。则

$$p(u_j) = \frac{q(u_j)}{Q}, \quad (11)$$

就是样本点落在 u_j 处的频率值, 可作为概率的估计值。那么超越 u_j 的概率值为

$$p(u_j) = \sum_{k=j}^m p(u_k), \quad (12)$$

$p(u_j)$ 即是超越概率风险估计值。

在信息扩散评估模型中, 扩散函数与扩散系数是关键, 直接关系到结果准确与否。本文选用的是最简单的正态扩散函数。黄崇福教授对不同扩散函数进行了验证^[10]。结果表明, 在样本容量不大的情况下, 简单正态分布要优于指数分布和对数正态分布。对于扩散系数, 许多研究者也对此做了探讨与改进^[11,12], 本文采用了黄崇福教授在《自然灾害风险评价——理论与实践》一书中所建议的扩散系数^[10]。

2 天津旱涝灾害指数

选取旱涝评估指标一般有两种思路: 一是从灾害影响后果出发, 选取旱涝受灾、成灾面积、粮食减产率、受灾人口等反映灾害后果的社会经济指标。这类指标反映了实际灾情, 但对灾害描述有一定滞后性, 而且在数据统计时易受人为因素影响致使获取的资料可能不准确。另一类指标, 是从旱涝诱发因素出发, 利用密切相关的降水等气象资料进行评估。气象资料具有及时性、科学性等特点, 而且不受人为因素干扰, 所以评估结果更为客观。本文选取降水距平百分率旱涝指数作为危险性评估的样本数据。

降水距平百分率反映了某时段降水量相对于同期平均状态的偏离程度。其具体计算公式(公式 13)和 7 级等级划分标准^[13](表 1)如下:

$$M_i = \frac{R_i - \bar{R}}{\bar{R}} \times 100\%, \quad (13)$$

式中： M_i 为降水距平百分率， i 表示年份； R_i 为某时段的降水量； \bar{R} 为同期降水量多年平均值。

表 1 降水距平百分率 7 级等级划分标准

等级	降水距平百分率 $M/\%$	旱涝等级
1	$M \geq 75$	重涝
2	$50 \leq M < 75$	大涝
3	$25 \leq M < 50$	偏涝
4	$-25 \leq M < 25$	正常
5	$-50 \leq M < -25$	偏旱
6	$-75 \leq M < -50$	大旱
7	$M \leq -75$	重旱

经计算，天津 6 ~ 9 月的降水量与全年降水量的相关系数约为 0.89，且约占全年降水量的 80%，所以这段时间的降水量能很好地反映降水量的年际变化。选取天津 1984 ~ 2005 年这 22 年的汛期降水资料为基础数据，通过计算每年的降水距平百分率得到危险性评估的旱涝灾害指数样本(表 2)。

表 2 1984 ~ 2005 年天津旱涝等级划分

年份	6 ~ 9 月降水量	旱涝 M_i 等等级划分	
	/mm	$M_i/\%$	旱涝类型
1984	464.3	12.249 6	正常
1985	484.5	17.133 2	正常
1986	462.0	11.693 5	正常
1987	391.0	-5.471 5	正常
1988	615.5	48.803 8	偏涝
1989	277.9	-32.814 7	偏旱
1990	405.2	-2.038 5	正常
1991	463.9	12.152 9	正常
1992	268.3	-35.135 6	偏旱
1993	377.7	-8.686 9	正常
1994	628.1	51.850 0	大涝
1995	592.1	43.146 6	偏涝
1996	403.1	-2.546 2	正常
1997	279.9	-32.331 1	偏旱
1998	294.4	-28.825 6	偏旱
1999	280.9	-32.089 4	偏旱
2000	351.6	-14.996 9	正常
2001	400.1	-3.271 5	正常
2002	338.0	-18.284 8	正常
2003	391.4	-5.374 8	正常
2004	366.7	-11.346 3	正常
2005	563.3	36.183 9	偏涝

3 天津旱涝危险性评估

在上面获得的 22 个灾害指数样本基础上，应用基于信息扩散理论的评估模型对天津旱涝危险

性进行评估。灾害指数样本集合为：

$$X = \{x_1, x_2, \dots, x_{22}\} = \{12.249\ 6, 17.133\ 2, 11.693\ 5, -5.471\ 5, 48.803\ 8, -32.814\ 7, -2.038\ 5, 12.152\ 9, -35.135\ 6, -8.686\ 9, 51.850\ 0, 43.146\ 6, -2.546\ 2, -32.331\ 1, -28.825\ 6, -32.089\ 4, -14.996\ 9, -3.271\ 5, -18.284\ 8, -5.374\ 8, -11.346\ 3, 36.183\ 9\}.$$

由降水距平百分率的计算公式知，样本指数的变化范围是 -100 ~ 100，所以灾害指数 x_i 的论域为 $[-100, 100]$ 。为了提高计算精度，选择步长为 5，将论域集合变成由 41 个点组成的离散集合。所以离散论域为：

$$U = \{u_1, u_2, \dots, u_{41}\} = \{-100, -95, \dots, 100\}$$

根据信息扩散风险评估模型，得到样本 X 落在 U 中各点 u_j 处的频率值(可作为概率估计值)，从而得到天津旱涝指数的概率密度曲线(图 1)。

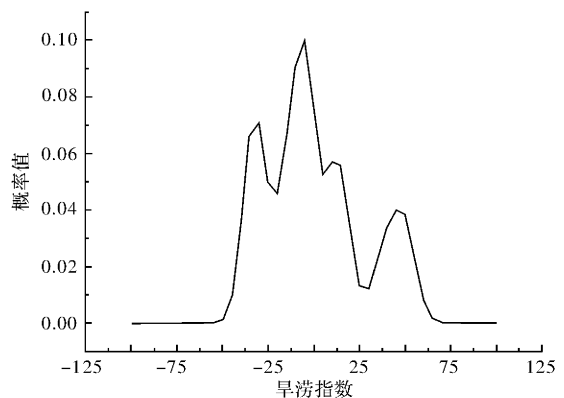


图 1 天津市旱涝指数概率分布曲线

根据概率计算公式 $P(u_m \leq u \leq u_n) = \sum_{k=m}^n p(u_k)$ ，我们得到不同等级旱涝危害危险性估计值(表 3)。

表 3 旱涝危险性估计值

灾害类型 (危害水平)	重旱	大旱	偏旱	正常	偏涝	大涝	重涝
危险性 估计值	0.000 0	0.001 3	0.231 3	0.638 0	0.159 1	0.073 4	0.000 0

从表 3 看出每个危害水平对应一个危险性评估值。例如，偏涝的估计值是 0.159 1，说明天津每年发生偏涝的可能性为 0.159 1，即约每 6 年会出现一次偏涝。同理，天津每年出现偏旱的可能性为 0.231 3，即约不到 5 年会出现一次偏旱。表 3 结果显示，重旱、重涝两种高强度灾害几乎不可能发生；大旱发生的可能性也不大，约 769 年一遇，但大涝则比较频发，约 14 年一遇。

表 4 是天津 1984 ~ 2005 年这 22 年各等级旱涝危害发生的次数及其频数。与其对比分析, 发现依据信息扩散理论评估模型得出的危险性评估结果(表 3)与这 22 年的实际情况基本吻合, 从而证明选用信息扩散理论评估模型对旱涝危险性评估是可行的。

表 4 天津市近 22 年各等级旱涝发生频率

等级	重旱	大旱	偏旱	正常	偏涝	大涝	重涝
发生的次数	0	0	5	13	3	1	0
频率	0.000 0	0.000 0	0.227 3	0.590 9	0.136 4	0.045 5	0.000 0

4 结论及讨论

在灾害危险性评价中往往很难得到准确全面的数据, 所以如何利用不完备信息研究尽可能准确的评价结果是危险性评估的关键。基于信息扩散理论的评估模型是一种优化利用不完备信息进行评估的技术方法。本文利用此模型对天津旱涝危险性进行评估并给出旱涝指数的概率分布曲线, 经验证, 评估结果与实际情况基本吻合。

从自然灾害系统论的角度看, 自然灾害是致灾因子、孕灾环境和承灾体三者相互作用的结果。自然灾害风险应是对这三者的反映, 应包括灾害事件发生的概率及其所导致的后果。因此, 基于降水量的旱涝指数概率只是对致灾因子发生可能性的描述, 本文所完成的只是旱涝灾害风险评价中的一步——危险性评价, 旱涝灾害的风险评价还需要做进一步的工作。

参考文献:

- [1] 谭桂容, 孙照勃, 陈海山. 旱涝指数的研究[J]. 南京气象学报, 2002, 25 (2): 153 - 158.
- [2] 陈家刚. 天津的旱涝灾害及其影响[J]. 天津师范大学报: 自然科学版, 1998, 18(4): 62 - 67.
- [3] 苏同卫, 李可军, 李岂秀, 等. 天津市及周围地区近 500 年旱涝变化分析[J]. 干旱气象, 2007, 25(1): 21 - 24.
- [4] 黄崇福, 王家鼎. 模糊信息优化处理技术及其应用[M]. 北京: 北京航空航天大学出版社, 1995.
- [5] 黄崇福, 张俊香, 刘静. 模糊信息优化处理技术应用简介[J]. 信息与控制, 2004, 33(1): 61 - 66.
- [6] 冯利华. 基于信息扩散理论的洪水风险分析[J]. 信息与控制, 2002, 31(2): 164 - 165.
- [7] 刘引鸽, 缪启龙, 高庆九. 基于信息扩散理论的气象灾害风险评价方法[J]. 气象科学, 2005, 25(1): 84 - 89.
- [8] 全仝, 王玲珍, 黄成敏. 基于信息扩散理论的云南省地震风险评估及管理研究[J]. 西北地震学报, 2006, 28(2): 180 - 183.
- [9] 吴息, 杨静. 利用信息扩散模式对浙江省大暴雨的风险分析[J]. 灾害学, 2002, 17(4): 7 - 10.
- [10] 黄崇福. 自然灾害风险评价: 理论与实践[M]. 北京: 科学出版社, 2005: 96 - 98.
- [11] 王新洲, 游扬声, 汤永净. 最优信息扩散估计理论及其应用[J]. 地理空间信息, 2003, 1(1): 10 - 21.
- [12] 李梅, 张洪波, 黄强, 等. 基于信息扩散估计的洪水风险分析[J]. 中北大学学报: 自然科学版, 2007, 28(3): 193 - 198.
- [13] 鞠笑生, 杨贤伟, 陈丽娟, 等. 我国单站旱涝指标确定和区域旱涝级别划分的研究[J]. 应用气象学报, 1997, 8(1): 26 - 33.

Drought and Waterlogging Hazard Assessment in Tianjin Based on Information Diffusion Model

Du Xiaoyan, Huang Suiliang and Zhao Qingxiang

(College of Environment Science and Engineering, Nankai University, Tianjin 300071, China)

Abstract: Using waterlogging indexes of recent 22 years as samples, the drought and waterlogging hazards in Tianjin are assessed by adopting the information diffusion theory based assessment model. The probability density curves of the drought and waterlogging indexes and probability of drought and waterlogging of different degrees in Tianjin are given. The results show that occurrence of severe drought and vast waterlogging is impossible in Tianjin. There is little possibility of occurrence of great drought in Tianjin, once in about 700 years. The partial drought, severe waterlogging and partial waterlogging occur frequently. Their annual possibilities are 0.2313, 0.1591 and 0.0734 respectively. This result basically coincides with the actual condition of Tianjin in recent 22 years. It indicates feasible to make risk assessment of natural disaster without enough information by using this model.

Key words: information diffusion; the anomalous percentage of precipitation; drought and waterlogging hazards; hazard assessment; Tianjin