

Statistical Significance of Typhoon Risk and Calculation of the Risk per Year

Chongfu Huang^{1,2,3}

¹Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education of China, Beijing Normal University, Beijing 100875, China

²State Key Laboratory of Earth Surface Processes and Resource Ecology (Beijing Normal University), Beijing 100875, China

³Academy of Disaster Reduction and Emergency Management, Ministry of Civil Affairs & Ministry of Education, the Peoples' Republic of China, Beijing 100875, China

台风风险的统计学意义及年度风险的计算

黄崇福^{1,2,3}

¹北京师范大学环境演变与自然灾害教育部重点实验室, 北京 100875, 中国

²地表过程与资源生态国家重点实验室(北京师范大学), 北京 100875, 中国

³民政部/教育部减灾与应急管理研究院, 北京 100875, 中国

Abstract

The risks in terms of statistical significance refer to ones that could be analyzed by using random samples. The typhoon risk in statistics is the average loss caused by typhoons. The risk per year calculated with a stochastic process has more clear physical meaning. Monte Carlo method can not improve the risk assessment. The information diffusion technique can improve estimation of risk under small-sample condition. Providing a platform to people living in typhoon-prone area to participate in risk assessment and timely update risk information, the Internet of intelligences could help us to improve estimation of typhoon risk.

Keywords: statistics, typhoon, risk, information diffusion, Internet of intelligences

摘要

具有统计学意义的风险,是指可用随机样本分析的风险。统计学意义上的台风风险,就是台风造成的损失

基金项目: 国家自然科学基金项目(编号: 41471424、41471426)。

平均值。用随机过程计算年度风险,物理意义更为清楚。蒙特卡罗法不能改善风险评估。在小样本条件下,信息扩散技术可改进风险的评估。用互联网为台风频发地区的民众提供参与风险评估的渠道,及时更新风险信息,有望提高风险分析结果的实用性。

关键词: 统计学; 台风; 风险; 信息扩散; 互联网

1. 引言

从数学的观点看,统计学是应用数学的一个分支,主要通过利用概率论建立的数学模型,对收集的大量数据和资料进行量化分析、总结,对相关现象进行推断和预测,为管理决策提供依据和参考。

从学科体系上看,统计学并非应用数学的一个分支,而是一门独立的学科。最早的统计学,是对各地区的历史、行政、科学、艺术、人口、资源和财富等社会和经济情况的比较、分析、具有社会科学特点。直到十七世纪,统计方法才与数学计算和推理方法开始结合,运用定量方法分析社会经济问题。今天,统计学家们研究的重点,仍然是数据收集、数据整理。数学模型的应用,只是统计学的一部分工作。正因为如此,西方的大学中,统计学并不设在数学系内。用数学方法进行统计分析的学科,通常称为数理统计。

当我们研究台风风险时，数据收集、数据整理也是非常重要的工作，但多由政府部门承担，我们更关注的是用什么样的数学模型处理获得的数据，才能较为可靠地评估出风险值。本文所指台风风险的统计学意义，是数理统计上的统计学意义^[1]。

2. 从定性的风险到定量的风险

风险是与某种不利事件有关的一种未来情景^[2]。从图1所示的2015-2016年中国房地产泡沫“魔鬼”，人们能感觉到崩盘的风险越来越大，但政府全力以赴保护泡沫，人们无法计算风险有多大，只能定性地认识到风险正在积累。

搬进新家的床没放好，会出现吃不饱睡不香的现象。风水学关于卧室内床的摆放，有许多的忌禁，其实是风险问题。图2所示的摆放均使睡在床上的人暴露身体受损的风险之中。

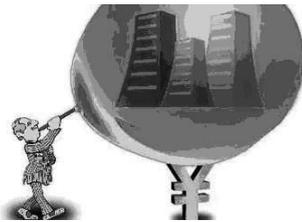


图1. 2015-2016年中国房地产泡沫“魔鬼”正加速累积着金融风险。

三月两月内，睡在摆放不当床上的人，身体受损的风险并不明显。长年累月睡在这种床上的人，风险不可忽视。收集整理大量与此相关的病例，进行统计

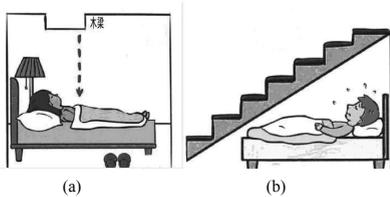


图2. 风水学关于卧室内床的摆放忌禁是风险问题。(a)睡在横梁下，人体的微电系统（尤其是大脑、心脏）会直接受到负面影响，造成睡眠不足、头痛等现象。(b)床在楼梯下方的人，睡眠状况一定不会好，经常恶梦不断，运势也会非常差。分析，就可用发病率的方式对床位摆放不当的风险进行量化。

在自然灾害风险分析领域^[3]，收集整理大量的历史灾害数据，进行系统的统计分析，定量地认识自然灾害风险，已经相当普及。

设某地区在过去的T年内某种自然灾害发生了n次，设第i次灾害事件的记录是 x_i 。我们将这些记录组成的集合 $X=\{x_1, x_2, \dots, x_n\}$ 称为样本，其内的元素 x_i 称为样本点。最简单的记录是损失；较全面的记录是致灾因子和灾情。最全面的记录可以穷尽一切，样本点中的一些分量甚至可以是文字等非量化内容。

我们可以用记录损失的样本对该种自然灾害的风险进行简单的量化分析：选用某种合适的统计方法，估计出X来自母体的概率分布 $P(x)$ ，其期望值 μ ，就是对损失风险的量化估计。这一估计值，其实就是 x_1, x_2, \dots, x_n 的平均值。换言之，平均损失值就是统计学意义上的损失风险值。

由于致灾因子的强度与灾害损失不是一对一的物理关系，自然灾害风险分析的基本模式^[4]，是估计致灾因子的概率分布和易损线曲线，由它们来计算损失期望值。我们用文献[5]中的20个台风灾害事件(表1)来展示两种分析结果。

表1. 2000-2006年影响浙江的20个台风

序号	名称	入境日期	级别	损失(亿)
1	启德	2000-7-10	11	6.8
2	杰拉华	2000-8-10	12	2.3
3	碧利斯	2000-8-23	12	2.3
4	派比安	2000-8-29	12	8
5	桑美	2000-9-12	12	31.4
6	飞燕	2001-6-23	9	0.8
7	威马逊	2002-7-3	12	7
8	森拉克	2002-9-7	13	46
9	环高	2003-8-22	9	0.7
10	蒲公英	2004-7-3	10	0.6
11	云娜	2004-8-12	12	181.3
12	艾利	2004-8-25	12	7.5
13	海马	2004-9-13	8	0.5
14	海棠	2005-7-19	12	92.9
15	麦沙	2005-8-6	12	89.1
16	泰利	2005-9-1	12	39.9
17	卡努	2005-9-11	15	87.2
18	珍珠	2006-5-18	10	1.2
19	碧利斯	2006-7-14	11	6.9
20	桑美	2006-10-17	17	127.3

以y和z分别记台风级别和损失（亿元），此容量为20的样本如式（1）所示。

$$X = \{(y_1, z_1), (y_2, z_2), \dots, (y_{20}, z_{20})\} \\ = \{(11, 6.8), (12, 2.3), (12, 2.3), (12, 8), (12, 31.4), (9, 0.8), (12, 7), (13, 46), (9, 0.7), (10, 0.6), (12, 181.3)\}$$

(12, 7.5), (8, 0.5), (12, 92.9), (12, 89.1), (12, 39.9), (15, 87.2), (10, 1.2), (11, 6.9), (17, 127.3)} (1)

由 z_1, z_2, \dots, z_{20} 计算得平均损失值是

$$\bar{z} = 36.985 \text{ (亿元)} \quad (2)$$

图 3 (a) 是 y_1, y_2, \dots, y_{20} 的频率直方图^[6]。用其估计的概率分布见图 3 (b)。

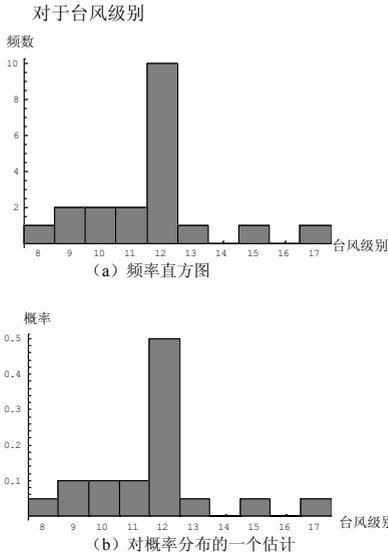


图 3. 用直方图方法, 依据台风的历史记录, 对不同级别台风发生的概率进行估计。例如, 由 (b) 知, 一旦台风发生, 有一半的可能性是 12 级台风。

$$m_1=8, m_2=9, \dots, m_{10}=17 \quad (3)$$

我们得到一个概率向量 L :

$$P=(p(m_1), p(m_2), \dots, p(m_{10})) \\ = (0.05, 0.1, 0.1, 0.1, 0.5, 0.05, 0, 0.05, 0, 0.05) \quad (4)$$

估计台风易损性最简单的方法, 是统计式 (1) 中不同级别台风的平均损失, 如图 4 所示。

对于式 (3) 中的台风级别, 我们得到一个平均损失向量 L :

$$L=(l(m_1), l(m_2), \dots, l(m_{10})) \\ = (0.5, 0.75, 0.9, 6.85, 46.17, 46, 0, 87.2, 0, 137.3) \quad (5)$$

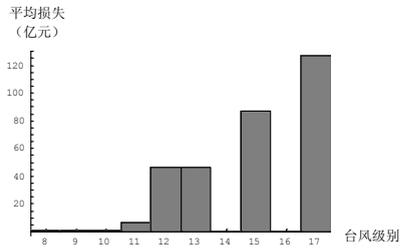


图 4. 由 20 个台风灾害事件统计出来的易损性关系。

由式 (4) 中概率向量 (分布) 和式 (5) 中损失向量, 估计出损失的期望值是:

$$\mu_{\text{Loss}}=0.5 \times 0.05+0.75 \times 0.1+0.90 \times 0.1+6.85 \times 0.1+46.17 \times 0.5+46 \times 0.05+0 \times 0+87.2 \times 0.05+0 \times 0+127.3 \times 0.05 \\ = 36.985 \text{ (亿元)} \quad (6)$$

即, 考虑到致灾因子发生概率和易损性估计出的损失期望值恰好是直接由表 1 中损失数据计算得到平均损失值。这不是巧合, 而是必然。

事实上, 设 y_i 级台风发生了 n_i 次, 损失分别是 $z_1(y_i), z_2(y_i), \dots, z_{n_i}(y_i)$ 。令 $i=1, 2, \dots, m$ 。于是平均损失值为:

$$\bar{z} = \frac{\sum_{1 \leq i \leq m} \sum_{1 \leq j \leq n_i} z_j(y_i)}{\sum_{1 \leq i \leq m} n_i} \quad (7)$$

易知, 由这些记录估计出 y_i 级台风发生的概率是

$$p(y_i) = \frac{n_i}{\sum_{1 \leq i \leq m} n_i} \quad (8)$$

y_i 级台风平均损失是

$$\bar{z}_i = \frac{\sum_{1 \leq j \leq n_i} z_j(y_i)}{n_i} \quad (9)$$

$\bar{z}_1, \bar{z}_2, \dots, \bar{z}_m$ 表达的, 就是由台风灾害记录估计出的易损性。于是, 损失的估计期望值为:

$$\mu_{\text{Loss}} = \sum_{i=1}^m \bar{z}_i p(y_i) \\ = \sum_{i=1}^m \frac{\sum_{1 \leq j \leq n_i} z_j(y_i)}{n_i} \frac{n_i}{\sum_{1 \leq i \leq m} n_i}$$

$$\begin{aligned}
 &= \sum_{i=1}^m \frac{\sum_{1 \leq j \leq n_i} z_j(y_i)}{\sum_{1 \leq i \leq m} n_i} \\
 &= \frac{\sum_{1 \leq i \leq m} \sum_{1 \leq j \leq n_i} z_j(y_i)}{\sum_{1 \leq i \leq m} n_i} \quad (10) \\
 &= \bar{z}
 \end{aligned}$$

式(10)说明,在统计学意义上,用自然灾害风险分析基本模式^[4]“致灾因子的概率密度函数和承灾体脆弱性曲线”计算台风风险是合理的。特别地,当历史灾害资料足够多,确保直方图估计出的概率分布和脆弱(易损)性曲线较可靠时,可直接用灾害的平均值作为风险值使用。然而,图3和图4说明,仅由表1的历史灾害资料,估计出的概率分布和易损性曲线并不可靠。不仅在14和16级台风处跳空,一些非跳空处也没有变化梯度,没有反映出应有的统计规律。

于是,为了算得“好看”一些,人们热衷于寻找理论分布来替代图3中的概率分布,寻找理论曲线来替代图4中的易损性关系,但统计特征必须与它们一致,估计出的损失期望值也应该基本相同。

如果统计资料充分多,理论上讲,用式(6)方式计算出的 μ_{loss} 就是风险值。然而,这一统计量只是说明:在统计时段内,发生一次台风,造成损失的最大可能性是36.985亿元。式(1)的样本,是七年记录,每年平均发生约3次台风灾害,是否 μ_{loss} 的3倍值就是该地区的年度风险呢?要回答这个问题,我们需要从随机过程的角度来观察台风灾害。

3. 用随机过程计算年度风险

事物的变化过程分为两大类:确定性过程和随机过程。确定性过程须满足下述两个条件:

- 1) 每次观察得到的过程都相同。
- 2) 观察到的过程可用时间 t 的确定函数表示。

随机过程则每次观察得到的过程都不同,观察到的过程也没有确定的变化形式,不能用一个时间 t 的确定函数表示。

例如,正弦信号就是一个确定性过程,而示波器的噪声电压就是一个随机过程。

我们以年度为观察台风灾害的时间长度。每一年观察到的各种台风发生时间都不一样,也不能用一个日期 t 的确定函数表示一年中台风的变化形式。

于是,针对表1的台风,我们可用分别记录各年台风发生情况的7个样本函数 $x_1(t), x_2(t), \dots, x_7(t)$ 来近

似描述台风的随机过程。注意到表1中台风发生的最短时间为6天,参考表1的日期,我们锁定台风季是从5月15日到10月20日。以5天为时间步长,每月按30天计,令:

$$\begin{cases} t_1 = [5月15日, 5月20日] \\ t_2 = [5月20日, 5月25日] \\ \dots \\ t_{31} = [10月15日, 10月20日] \end{cases} \quad (11)$$

于是,由表1可得到近似描述台风的随机过程的7个样本函数,列入表2。将表中的台风级数替换成式(5)中相应的平均损失,我们得到近似描述浙江台风灾害随机过程的样本函数,列入表3中。

表2. 近似描述影响浙江的台风随机过程的样本函数

	$x_1(t)$	$x_2(t)$	$x_3(t)$	$x_4(t)$	$x_5(t)$	$x_6(t)$	$x_7(t)$
t_1	0	0	0	0	0	0	10
t_2	0	0	0	0	0	0	0
t_3	0	0	0	0	0	0	0
t_4	0	0	0	0	0	0	0
t_5	0	0	0	0	0	0	0
t_6	0	0	0	0	0	0	0
t_7	0	0	0	0	0	0	0
t_8	0	9	0	0	0	0	0
t_9	0	0	0	0	0	0	0
t_{10}	0	0	12	0	10	0	0
t_{11}	11	0	0	0	0	0	0
t_{12}	0	0	0	0	0	0	11
t_{13}	0	0	0	0	0	12	0
t_{14}	0	0	0	0	0	0	0
t_{15}	0	0	0	0	0	0	0
t_{16}	0	0	0	0	0	0	0
t_{17}	12	0	0	0	0	12	0
t_{18}	0	0	0	0	12	0	0
t_{19}	0	0	0	0	0	0	0
t_{20}	12	0	0	9	12	0	0
t_{21}	12	0	0	0	0	0	0
t_{22}	0	0	0	0	0	12	0
t_{23}	0	0	13	0	0	0	0
t_{24}	12	0	0	0	8	15	0
t_{25}	0	0	0	0	0	0	0
t_{26}	0	0	0	0	0	0	0
t_{27}	0	0	0	0	0	0	0
t_{28}	0	0	0	0	0	0	0
t_{29}	0	0	0	0	0	0	0
t_{30}	0	0	0	0	0	0	0
t_{31}	0	0	0	0	0	0	17

随机过程的数字特征是指先把时间 t 固定,然后用随机变量的分析方法计算的数字特征。随机过程的数字特征是时间的函数。用样本函数估计的随机过程的数字特征期望,就是各样本函数在时间点上

的统计均值。表 3 的最后一列是其近似描述的随机过程的期望估计。

表 3. 近似描述浙江台风灾害随机过程的样本函数

	$x_1(t)$	$x_2(t)$	$x_3(t)$	$x_4(t)$	$x_5(t)$	$x_6(t)$	$x_7(t)$	期望
t_1	0	0	0	0	0	0	0.9	0.1286
t_2	0	0	0	0	0	0	0	0
t_3	0	0	0	0	0	0	0	0
t_4	0	0	0	0	0	0	0	0
t_5	0	0	0	0	0	0	0	0
t_6	0	0	0	0	0	0	0	0
t_7	0	0	0	0	0	0	0	0
t_8	0	0.75	0	0	0	0	0	0.1071
t_9	0	0	0	0	0	0	0	0
t_{10}	0	0	46.17	0	0.9	0	0	6.7243
t_{11}	6.85	0	0	0	0	0	0	0.9786
t_{12}	0	0	0	0	0	0	6.85	0.9786
t_{13}	0	0	0	0	0	46.17	0	6.5957
t_{14}	0	0	0	0	0	0	0	0
t_{15}	0	0	0	0	0	0	0	0
t_{16}	0	0	0	0	0	0	0	0
t_{17}	46.17	0	0	0	0	46.17	0	13.1914
t_{18}	0	0	0	0	46.17	0	0	6.5957
t_{19}	0	0	0	0	0	0	0	0
t_{20}	46.17	0	0	0.75	46.17	0	0	13.2986
t_{21}	46.17	0	0	0	0	0	0	6.5957
t_{22}	0	0	0	0	0	46.17	0	6.5957
t_{23}	0	0	46	0	0	0	0	6.5714
t_{24}	46.17	0	0	0	0.5	87.2	0	19.1243
t_{25}	0	0	0	0	0	0	0	0
t_{26}	0	0	0	0	0	0	0	0
t_{27}	0	0	0	0	0	0	0	0
t_{28}	0	0	0	0	0	0	0	0
t_{29}	0	0	0	0	0	0	0	0
t_{30}	0	0	0	0	0	0	0	0
t_{31}	0	0	0	0	0	0	137.3	19.6143

台风灾害随机过程时间段 t 上的期望值，其物理意义就是此时段的台风风险。显然，一年的台风风险就是所有时段的台风风险求和。对表 3 中各时段期望求和，我们得到年度风险：

$$R=0.1286+0.1071+\dots+19.6143=107.1(\text{亿元}) \quad (12)$$

表 3 中所有样本函数的数值求和再除以 7，可直接得出这一数值。

7 年中 20 个台风记录，平均每年有 2.8571 个台风。以此数直接乘式 (2) 中的平均损失值，我们得：

$$2.8571 \times 36.985 = 105.6698 \text{ (亿元)}$$

它与式 (12) 中的年度风险值 R 近似相等。可见，用 $t_{L_{\text{loss}}}$ 的 3 倍值 (110.955) 作为是该地区的年度风险估计值，误差不大，也是可以的。

4. 用信息扩散技术对简单统计风险进行改进

从现实角度出发，无论是笼统的风险分析，还是用随机过程计算年度风险，因为数据太少，由表 1 进行的简单统计，结果的可靠性都令人怀疑。用很少的记录简单统计的风险，很不靠谱。

于是，一些研究人员用蒙特卡罗法来研究台风风险，可以生成任意多的台风样本。由于巧妙的包装，许多学者不明就里，以为用蒙特卡罗法帮助，可以更好地估计风险。其实不然。

蒙特卡罗法亦称作随机模拟方法^[4]。任何一种能够利用随机数实现模拟的方法，都可称为蒙特卡罗法^{[7][8]}。该方法能帮助人们直观地观察随机数的行为，提高对随机系统的认识。然而，任何蒙特卡罗方法的实现，都必须由一个或几个概率函数控制的伪随机发生器进行抽样来实现。换言之，在台风风险分析中使用蒙特卡罗法，先决条件是，用于产生台风样本的概率函数能表达台风的随机规律。既然相关概率函数已知，何必多此一举又用抽取的样本来估计风险？诚如本文式 (6) 的 $\mu_{L_{\text{loss}}}$ 是式 (2) 中的平均损失值的虚晃一枪一样，假定概率分布后又用其生成的样本来估计此概率分布，也是虚晃一枪。也就是说，如果控制抽样的概率函数不能近似描述台风的真实随机规律，由蒙特卡罗法计算的台风风险值更不靠谱，甚至不如真实的小样本简单统计出来的结果。

在小样本条件下，我们建议用信息扩散技术对简单统计结果进行改进。

目前，信息扩散理论的基础比较稳固，应用涉及面较广。信息扩散的最初思想是在研究震害面积和震级间的关系考虑如何使少量地震记录发挥更大作用时产生的。严格定义是在 1997 年的文献[9]中给出。随着研究的不断深入，建立了六个实用的模型体系：(1)用信息分配方法计算软直方图；(2)用正态信息扩散估计概率分布；(3)生成输入-输出型模糊关系矩阵的自学习离散回归模型；(4)能处理矛盾样本的信息扩散人工神经网络；(5)计算机仿真输入非参数化模型；(6)计算模糊风险的内集—外集模型。

随后，王家鼎、王新洲、沈赞琦、Moraga、Karimi、Makó 等一大批人围绕信息扩散理论和方法做了大量的工作。王家鼎 (1999) 通过求 Laplace 方程提出了一个对数型扩散函数，用于研究黄土斜坡稳定性问题，计算值与实际值的符合程度很高^[10]；王新洲等 (2003) 用渐近无偏准则给出了一个计算最优正态扩散系统的方法，用于处理轨道交通工程的水准测量数据，具有良好的抗差性^[11]；沈赞琦等 (2007)

Risk Analysis and Crisis Response in Big Data Era (RAC-16)

用 MacCormack 方法求解非均匀各向异性体扩散方程提出了一个二维非均匀信息扩散, 用于研究保险产品 中高血压二维患病率的问题, 减少误差约 70%^[12]; Moraga 等 (2003) 根据算法复杂度理论, 将内集—外集模型的组合算法, 简单化为排序算法, 彻底解决了计算困难的问题^[13]; Makó (2005) 用分析几何理论证明了信息扩散原理的普适性^[14]; Karimi 等 (2007) 综合运用贝叶斯理论和概率密度向可能性转化的方法, 发展出的一个内集—外集模型的贝叶斯算法, 并用于评估伊斯坦布尔地区的地震风险, 证明了方法的实用性^[15]。

信息扩散理论和方法在风险分析方面的应用较多。从近年来在 Elsevier 出版刊物上发表的文章看, 主要有, 用于分析太湖蓝藻暴发风险, 为政府决策提供了重要依据^[16]; 用于研究美国东海岸飓风风险, 计算出保守风险值和冒险风险值^[17]; 用于研究中国最近 20 年的水灾数据, 计算出不同可能性的多值风险^[18]; 用于研究中国北方草原火灾风险, 为畜牧业生产制定补偿计划提供了依据^[19]; 用于研究洞庭湖区洪水、干旱、虫害和鼠害, 绘制出粮食生产的自然灾害风险图^[20]; 用正态扩散处理我国水利部数据库中的水灾数据, 生成可变模糊集, 拟合出了水灾风险曲线^[21]; 用于北塞浦路斯旅游保险气候指数计算, 绘制了风险图^[22]; 将信息扩散模型嵌入土壤流失方程, 评估环渤海地区不同降雨情况下的土壤侵蚀风险^[23]; 用于分析液化天然气站过去 12 个月的运行数据, 能及时发现潜在的风险^[24]; 用于研究草原生物灾害, 绘制出我国北方 10 省区的风险图^[25]; 用于研究农业保险业务数据, 计算不同损失的可能性, 为政府财政支持农险提供依据^[26]; 用于广东省北江、西江和绥江三江汇流区的洪水风险评估, 能根据洪峰水位推断水灾面积的几种可能性^[27]; 用于广州南沙区黄阁镇和南沙镇的区域环境风险评估, 帮助当地政府优化工业区布局, 建立风险防范管理程序^[28]。

我们用正态信息扩散估计概率分布的模型^[4]来处理式 (1) 中 y_1, y_2, \dots, y_{20} 的台风数据, 得图 5 所示的概率分布估计。此估计明显比图 3(b) 更具说明力, 因为它消除了平缓 and 跳空, 表现出一定的统计规律。

使用正态信息扩散方法构造信息矩阵^[29], 再用全条件落影公式进行推理^[30], 识别台风与损失间的关系, 得图 6 所示的易损性曲线。不同级别台风造成的损失列入表 4。此易损性曲线明显比图 4 的易损性关系合理, 因为现实中 14 级和 16 级的台风一定会造成相当大的损失, 而不是图 4 中的损失为 0。

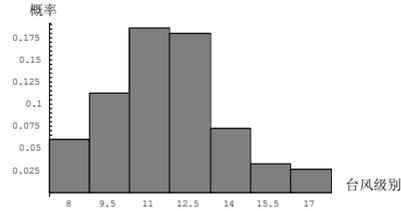


图 5. 用正态信息扩散方法, 依据台风的历史记录, 对不同级别台风发生的概率进行估计。

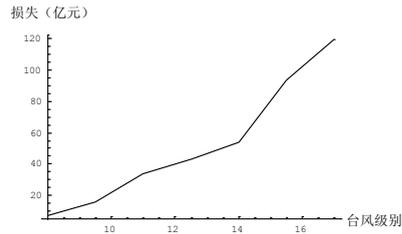


图 6. 用信息扩散技术处理 20 个台风灾害事件估计出的易损性曲线。

表 4. 用信息扩散技术处理 20 个台风灾害事件估计出不同级别台风造成的损失。

台风级别	造成的损失 (亿元)
8	7.535
9	13.001
10	21.694
11	33.614
12	40.213
13	47.008
14	54.001
15	80.130
16	101.991
17	119.585

由图 5 的概率分布和图 6 的易损性曲线估计的损失期望值为:

$$\tilde{\mu}_{\text{Loss}} = 39.462 \quad (13)$$

它比简单统计出的式(6)中的风险 μ_{Loss} 略大。研究证明, 信息扩散技术估计出的风险较为稳定和可靠。

用表 4 中台风灾害损失数据替代表 3 中相应台风级别台风损失, 得到优化处理小样本后近似描述浙江台风灾害随机过程的样本函数和相应的期望, 列于表 5 中。对表中各时段期望求和, 我们得到年度风险:

$$R=3.0991+1.8573+\dots+17.0836=113.2866 \text{ (亿元)} \quad (14)$$

它比简单统计的,由表 3 进行估计,在式(5)中给出的年度风险值 107.1 亿元略大。

表 5. 优化处理后的台风灾害随机过程的样本函数

	$x_1(t)$	$x_2(t)$	$x_3(t)$	$x_4(t)$	$x_5(t)$	$x_6(t)$	$x_7(t)$	期望
f_1	0	0	0	0	0	0	21.694	3.0991
f_2	0	0	0	0	0	0	0	0.0000
f_3	0	0	0	0	0	0	0	0.0000
f_4	0	0	0	0	0	0	0	0.0000
f_5	0	0	0	0	0	0	0	0.0000
f_6	0	0	0	0	0	0	0	0.0000
f_7	0	0	0	0	0	0	0	0.0000
f_8	0	13.001	0	0	0	0	0	1.8573
f_9	0	0	0	0	0	0	0	0.0000
f_{10}	0	0	40.213	0	21.694	0	0	8.8439
f_{11}	33.614	0	0	0	0	0	0	4.8020
f_{12}	0	0	0	0	0	0	33.614	4.8020
f_{13}	0	0	0	0	0	40.213	0	5.7447
f_{14}	0	0	0	0	0	0	0	0.0000
f_{15}	0	0	0	0	0	0	0	0.0000
f_{16}	0	0	0	0	0	0	0	0.0000
f_{17}	40.213	0	0	0	0	40.213	0	11.4894
f_{18}	0	0	0	0	40.213	0	0	5.7447
f_{19}	0	0	0	0	0	0	0	0.0000
f_{20}	40.213	0	0	13.001	40.213	0	0	13.3467
f_{21}	40.213	0	0	0	0	0	0	5.7447
f_{22}	0	0	0	0	0	40.213	0	5.7447
f_{23}	0	0	47.008	0	0	0	0	6.7154
f_{24}	40.213	0	0	0	7.535	80.130	0	18.2683
f_{25}	0	0	0	0	0	0	0	0.0000
f_{26}	0	0	0	0	0	0	0	0.0000
f_{27}	0	0	0	0	0	0	0	0.0000
f_{28}	0	0	0	0	0	0	0	0.0000
f_{29}	0	0	0	0	0	0	0	0.0000
f_{30}	0	0	0	0	0	0	0	0.0000
f_{31}	0	0	0	0	0	0	119.585	17.0836

5. 用互联网技术提高风险分析结果的实用性

通常,一个地区几十年内发生的台风,规律没有明显的变化。因为许多研究表明,要么变化不大;要么关于频数的结论相反。例如, Kamahori 等利用来自日本东京台风中心(JMA)的和联合台风警报中心(JTWC)的最佳路径热带气旋资料,对西北太平洋 1977-2004 年间台风天数进行对比分析,发现 JMA 资料显示的 4、5 级台风天数从 1977-1990 年间每年 7.2 天减少到 1991-2004 年的每年 4.3 天;而 JTWC 资料显示的 4、5 级台风天数从 1977-1990 年间的每年 9.8 天增加到 1991-2004 年间每年 16.9 天^[31]。一些全球气候变化模式模拟结果表明全球台风频数减少^[32];另外一些模式模拟的结果却显示台风频数增加^[33]。

然而,一个地区几十年内社会经济系统通常会发生明显变化,台风风险必然随之变化。此种情况下,依据历史灾害资料估计的风险,无论是简单统计的结果,还是信息扩散技术得来估计值,都很不准确,实用性不高。人们必须根据系统的变化调整风险估计值,提高实用性。

目前,社会经济系统变化跟踪往往滞后较长时间,从而大大降低了调整风险估计值的时效性。台风频发地区的民众,是台风风险的利益相关者,但不能实质性地参与风险评估工作,难以提高风险估计值的时效性。互联网的出现,为解决这一问题提供了一个选择。

互联网的最初思想是在研究如何进行风险分析在线服务时产生的^[34]。通俗地讲,由互联网联结具有观察、演绎、推理和解决问题能力的多个人(智能体),并通过嵌入的模型集个体小智慧为群体大智慧的网络,称为互联网。严格定义是在 2015 年的文献[35]中给出,即,设 A 是一个智能体集合, N 是 A 使用的一个网络, M 是处理 A 所提供信息的模型,三元体 $\langle A, N, M \rangle$ 称为一个互联网。温州地区水产养殖业台风灾害的在线研究,验证了互联网的有效性^[36];北京某社区风险雷达的实践,展示了互联网驱动风险雷达的能力^[37]。

网络让数据长了翅膀,智慧让信息有了灵魂,人机结合的互联网有助于推动风险分析进入互联网的大数据时代:自然灾害信息将实现高效传输、广泛覆盖和精准推送;风险承受体的信息将实现全面表达、交叉验证和及时更新;利益相关者将有充分的机会分享信息、贡献知识和提出判断。这一切,既有助于获得更多的数据资料,更有助于获得宝贵的利益相关者经验,有效提高灾害风险评估的质量。

用互联网技术提高台风风险分析结果实用性的基本原理是,由台风频发地区的民众和相关专家组成智能体集合 A ,以互联网作为 A 使用的网络 N ,根据 A 可能提供的柔性信息和参与程度,由专家配置相应的模型库 M ,充分发挥经验、数据库、监控系统、判断、推理、组合、修正和风险分析模型等作用,依据过去、当前和未来可能的发展,评估出未来一定时段内指定地区台风灾害的期望数值。显然,互联网的主要作用在于由民众来细化并及时更新风险信息,经由处理非结构化信息的复式互联网^[38],认知真正的风险,而非用过去的的数据,认识灾害系统过去的表现。

6. 结论与讨论

理论上讲,自然灾害风险是简单的风险问题^[39]。

频繁且随机发生的台风，其风险是概率风险。统计学意义上的台风风险，是大量台风灾害的平均值；概率意义上讲，就是损失的期望值。

考虑到致灾因子发生概率和易损性估计出的损失期望值，必然等于直接由损失数据计算得到平均损失值。当统计资料较少时，属于小样本问题，传统统计方法估计出的风险值较为粗糙，可用信息扩散技术来提高评估精度。

可以将台风灾害标于一年的各时间段上，多年的记录形成多个样本函数，近似描述台风灾害的随机过程。各样本函数在时间段上的统计均值，是此时间段上的风险。累积一年的各时间段的风险，可计算出台风年度风险，它与平均损失值乘以年平均台风发生数所得结果，差别很小。

在台风风险分析中使用蒙特卡罗法，必须知道台风随机规律的概率密度函数，用其来控制伪随机数发生器。使用须要估计的概率分布去控制伪随机数发生器，抽取伪随机数再来估计此概率分布，在风险分析中毫无意义。可见，蒙特卡罗法并不能改善风险评估。

在“人人都是传感器”的网络时代，人机结合的智联网，通过“网络让数据长了翅膀，智慧让信息有了灵魂”，正在推动风险分析进入互联网的大数据时代。

参考文献

- [1] D. Marcek, L. Falat, Volatility forecasting in financial risk management with statistical models and ARCH-RBF neural networks, *Journal of Risk Analysis and Crisis Response*, 4(2): 77-95, 2014.
- [2] C.F. Huang and D. Ruan, Fuzzy risks and an updating algorithm with new observations, *Risk Analysis*, 28(3): 681-694, 2008.
- [3] X.L. Pang, Dynamic vulnerability analysis of population for flood disaster in urban area, *Journal of Risk Analysis and Crisis Response*, 3(4): 166-174, 2013.
- [4] 黄崇福, 自然灾害风险分析与管理, 科学出版社, 北京, 2012 年.
- [5] 黄崇福, 自然灾害动态风险分析的一个虚拟案例, *灾害学*, 30(4): 1-11, 2015.
- [6] P. Bakhshi, A. Touran, A new approach for contingency determination in a portfolio of construction projects, *Journal of Risk Analysis and Crisis Response*, 2(4): 223-232, 2012.
- [7] M.Y. Li, D.Y. Huang, M. Liu, MCSim-based occupational health risk assessment on benzene, *Journal of Risk Analysis and Crisis Response*, 3(3): 135-145, 2013.
- [8] X.K. Zeng, D. Wang, J.H. Wu, Evaluating the three methods of goodness of fit test for frequency analysis, *Journal of Risk Analysis and Crisis Response*, 5(3), 178-187.
- [9] C.F. Huang, Principle of information diffusion, *Fuzzy Sets and Systems*, 91(1): 69-90, 1997.
- [10] 王家鼎, 地理学研究中的模糊信息优化处理方法, *地理学与国土研究*, 15(1): 75-80, 1999.
- [11] 王新洲, 游扬声, 汤永净, 最优信息扩散估计理论及其应用, *地理空间信息*, 1(1): 10-17, 2003.
- [12] 沈赞琦, 陆余楚, 二维非均匀信息扩散及其应用, *应用数学与计算数学学报*, 21(2): 91-96, 2007.
- [13] C. Moraga, C.F. Huang, Learning subjective probabilities from a small data set, *Proceedings of 33rd International Symposium on Multiple-Valued Log* (Tokyo, May 16-19, 2003), pp. 355-360.
- [14] Z. Makó, Approximation with diffusion-neural-network Presented in 6th International Symposium of Hungarian Researchers on Computational Intelligence (Budape: Hungary, November 18-19, 2005).
- [15] I. Karimi, E. Hüllermeier, Risk assessment system - natural hazards: A new approach based on fuzzy probability, *Fuzzy Sets and Systems*, 158(9): 987-999, 2007.
- [16] Q. Chen, H. Rui, W. Li, Y. Zhang, Analysis of algal bloom risk with uncertainties in lakes by integrating self-organizing map and fuzzy information theory, *Science of the Total Environment*, 482-483(1): 318-322, 2014.
- [17] L. Feng, G. Luo, Application of possibility-probability distribution in risk analysis of landfall hurricane—A case study along the east coast of the United States, *Applied Soft Computing*, 11 (8): 4563-4568, 2011.
- [18] Q. Zou, J. Zhou, C. Zhou, J. Guo, W. Deng, M. Yang, Liao, Fuzzy risk analysis of flood disasters based on a diffused-interior-outer-set model, *Expert Systems with Applications*, 39 (6): 6213-6220, 2012.
- [19] X. Liu, J. Zhang, W. Cai, Z. Tong, Informatic diffusion-based spatio-temporal risk analysis of grassland fire disaster in northern China, *Knowledge-Based Systems*, 23 (1): 53-60, 2010.
- [20] L. Zhong, L. Liu, Y. Liu, Natural disaster risk assessment of grain production in Dongting Lake Area, China, *Agriculture and Agricultural Science Proceedings*, 1(1): 24-32, 2010.
- [21] Q. Li, J. Zhou, D. Liu, X. Jiang, Research on flood risk analysis and evaluation method based on variable fuzzy sets and information diffusion, *Safety Science*, 50(1): 1275-1283, 2012.
- [22] H.G.T. Olya, H. Alipour, Risk assessment of precipitation and the tourism climate index, *Tourism Management*, 50: 73-80, 2015.
- [23] L. Xu, X. Xu, X. Meng, Risk assessment of soil erosion in different rainfall scenarios by RUSLE model coupled with information diffusion model: A case study of Bohai Rim, China, *Catena*, 100(2): 74-82, 2012.
- [24] Y.Y. Chu, W.L. Dong, Y. Li, D. Liang, Risk prediction model of LNG terminal station based on informatic diffusion theory, *Procedia Engineering*, 52: 60-66, 2012.
- [25] L. Hao, L. Yang, J.M. Gao, The application of information diffusion technique in probabilistic analysis of grassland biological disasters risk, *Ecological Modelling*, 272: 264-270, 2014.
- [26] L. Xing, K. Lu, The importance of public-private partnerships in agricultural insurance in China: based on analysis for Beijing, *Agriculture and Agricultural Science Proceedings*, 1(1): 241-250, 2010.
- [27] Q. Zou, J. Zhou, C. Zhou, L. Song, J. Guo, Y. Liu, TI practical research on flood risk analysis based on IIOSI and fuzzy α -cut technique, *Applied Mathematics*

- Modelling*, 36(7): 3271-3282, 2012.
- [28] L. Xu, G. Liu, The study of a method of regional environmental risk assessment, *Journal of Environmental Management*, 90(11): 3290-3296, 2009.
- [29] 黄崇福, 自然灾害风险分析的信息矩阵方法, *自然灾害学报*, 15(1):1-10, 2006.
- [30] 黄崇福, *自然灾害风险评价: 理论与实践*, 科学出版社, 北京, 2005年.
- [31] H. Kamahori, N. Yamazaki, N. Mannoji and K. Takahashi, Variability in intense tropical cyclone days in the western North Pacific, *SOLA*, 2: 104-107, 2006.
- [32] M. Sugi, A. Noda and N. Sato, Influence of global warming on tropical cyclone climatology: an experiment with the JMA global model, *Journal of the Meteorological Society of Japan*, 80(3): 249-272, 2002.
- [33] R.E. McDonald, D.G. Bleaken, D.R. Cresswell, V.D. Pope and C.A. Senior, Tropical storms: representation and diagnosis in climate models and the impacts of climate change, *Climate Dynamics*, 25(1):19-36, 2006.
- [34] C.F. Huang, Internet of intelligences in risk analysis for online services, *Journal of Risk Analysis and Crisis Response*, 1(2): 110-117, 2011.
- [35] C.F. Huang, Internet of intelligences can be a platform for risk analysis and management, *Human and Ecological Risk Assessment*, 21(5): 1395-1409, 2015.
- [36] 艾福利, *自然灾害风险分析互联网服务平台构建与应用研究*, 北京师范大学博士学位论文, 2013..
- [37] 吴彤, *互联网驱动的在线风险雷达在社区应急管理中的应用研究*, 北京师范大学硕士学位论文, 2015.
- [38] C.F. Huang, Multiple Internet of intelligences for risk analysis, *Journal of Risk Analysis and Crisis Response*, 4(2): 61-71, 2014.
- [39] IRGC, *Risk Governance - Towards an Integrative Approach*, White Paper No.1. Geneva: International Risk Governance Council, 2005.