

# 北京市感染性腹泻疾病的医疗 气象预报与应用研究

张德山<sup>1</sup> 孙培源<sup>2</sup> 赵 娜<sup>1</sup> 蒲永兰<sup>2</sup>  
韦懿云<sup>2</sup> 宛公展<sup>3</sup> 张姝丽<sup>1</sup>

(1. 北京市专业气象台, 100089; 2. 北京市海淀区疾病预防控制中心;  
3. 天津市气象科学研究所)

**提 要:**通过分析北京市海淀区感染性腹泻病发病情况与北京市观象台地面气象要素的相关关系,建立感染性腹泻病医疗气象分级预报及风险水平评估模式。收集海淀区2004—2006年逐日菌痢周发病数和同期气象因素资料,建立数据库,分析气象因素与肠道传染病发病率的关系,并采用多元回归概率分级技术,建立感染性腹泻病医疗气象分级预报(划分5个等级)及风险水平评估模式。结果表明,感染性腹泻病发病与水汽压的相关程度很好,单相关系数为0.8340。在所有候选预报因子中,它的方差贡献占70.5%。其他方差贡献较大的因子,还有平均风速、气温日较差等。在感染性腹泻病医疗气象分级预报及风险水平评估模式中,如果预报等级与实际等级相差 $\leq \pm 1$ 级统计为正确,感染性腹泻病医疗气象分级预报模式组的平均预报正确率可达98.5%。其中,预报等级与实际等级完全吻合的正确率可到62.9%~66.4%(指6个预报时效)。该模式预报结果可较好地为2008年奥运会提供肠道传染病预防服务,为北京各级疾病预防控制中心指导广大城乡居民卫生防疫提供科学依据。

**关键词:** 感染性腹泻 疾病流行风险水平评估 医疗气象预报

## Research on the Application of Medical-Meteorological Forecast Model of Infectious Diarrhoea Disease in Beijing

Zhang Deshan<sup>1</sup> Sun Peiyun<sup>2</sup> Zhao Na<sup>1</sup> Pu Yonglan<sup>2</sup> Wei Yiyun<sup>2</sup>  
Wan Gongzhan<sup>3</sup> Zhang Shuli<sup>1</sup>

(1. Beijing Professional Weather Service Observatory, 100089;  
2. Center for Disease Prevention and Control of Haidian District, Beijing; 3. Weather scientific Academe of Tianjin)

资助项目: 2008年奥运会传染病防治保障项目“气象因素与腹泻病发病关系及预测研究”;项目来源:北京市疾病预防控制中心2006年度传染病预防控制项目。

收稿日期: 2008年4月9日; 修定稿日期: 2008年6月27日

**Abstract:** Based on the data, collected in Haidian district in 2006, of the incidence of pre-day's infectious diarrhoea and the weather condition, a forecasting model of infectious diarrhea medical-meteorological grade and risk level evaluation model were set up by using multiple regression probability grade analysis. The Pearson correlation coefficient of the incidence and the atmospheric vapor pressure was 0.8340. The proportion of variance contribution of vapor pressure was highest (70.5%). Mean wind speed and diurnal temperature had higher proportion of variance contribution than other factors. If the difference between the forecasted grade and actual grade is no more than 1, the forecasted grade of the infectious diarrhoea disease medical-meteorological grade forecast models would be defined as proper grade. The mean ratio of proper forecasted grade forecast models was 98.5%, and the exact proper rate (for 6 time periods) was 62.9~66.4%. It is concluded that the forecasting model of infectious diarrhoea disease medical-meteorological grade forecast and the risk level evaluation model could well forecast the incident of infectious disease during Beijing Olympic Games, and be helpful for the Center of Disease Prevention and Control (CDC) to provide public health services for citizens according to the outcomes of the models.

**Key Words:** infectious diarrhoea disease epidemic risk level evaluation medical-meteorological forecast

## 引 言

感染性腹泻疾病是北京地区夏、秋季节的常见流行性疾病或疫情。它的发生、发展与蔓延与环境气象条件密切相关。根据已有的历史病例和同期天气气候资料,研究感染性腹泻疾病与天气气候的相关关系,建立医疗气象预报模型,可逐日滚动预测未来 6 种时效(即 24~48、24~72、24~96、24~120、24~144、24~168 小时,下文不再详述这 6 个时效)的疾病疫情发生、发展与蔓延趋势及风险水平评估产品,有利于贯彻“预防为主”的方针,可在很大程度上为疾病预防控制中心提供疾病可能流行的气候背景,指导广大城乡居民防病治病,保障居民身体健康<sup>[1-2]</sup>。

## 1 资料与因子相关分析

### 1.1 资料来源

感染性腹泻疾病逐日发病病例,取自北

京市海淀区疾病预防控制中心。自 2004 年 1 月 1 日至 2006 年 12 月 31 日,共搜集到病例样本 1092 个,仅有 4 天缺测。病例涵盖人口,约为 250 万。同步地面气象因子由北京市气象台提供,包括逐日地面基本气象要素 9 个(含 1. 平均气温 2. 极端最高气温 3. 极端最低气温 4. 平均相对湿度 5. 最小相对湿度 6. 平均水汽压 7. 日降水量 8. 平均风速 9. 平均海平面气压)。因子样本数  $9 \times 1096$  个,没有缺测。

### 1.2 感染性腹泻疾病发病的统计学特征

感染性腹泻疾病,一年四季都有发生,但以盛夏至秋季最甚,此现象与北京的暴雨季相吻合<sup>[3]</sup>,详见图 1。感染性腹泻疾病发病的统计参数值,详见表 1。从表 1 所列数据不难看出,感染性腹泻疾病日平均发病人次,每百万人口约为 7.2,中位数为 4.4,由于中位数远小于平均值,概率分布显然不属于正态分布类型。

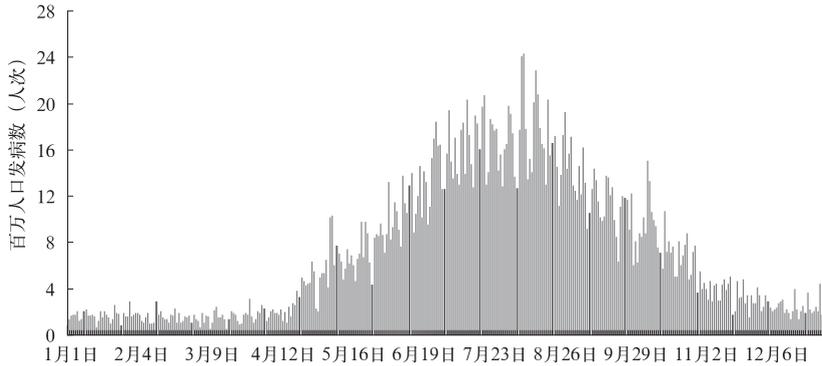


图 1 2004—2006 年感染性腹泻疾病平均发病人数

表 1 北京市海淀区感染性腹泻疾病逐日发病的统计特征值(单位:人次/每百万人口)

特征分类	最大值	次大值	第三名	最小值	中位数	平均值	标准差
统计参数	51.14481	41.58115	37.42303	0.39984	4.4	7.2	7.1
出现日期	2006-08-09	2006-08-08	2006-08-14	2004-01-25			

### 1.3 感染性腹泻疾病发病与环境气象因子的相关分析

所谓环境气象因子,是指根据北京市气象台 9 个地面基本气象要素,按一定方法加工派生而成<sup>[4-5]</sup>。其中,因子 1、4、5、6、7、8、9 是因子源值的等权或半衰权重的平均值;因子 2 和 10 是过去 7 天逐日的气温、气压变化倾向值,其计算方法是将因子源值先用直线方程拟合,求解得出斜率值再乘以方程相关系数的绝对值;因子 3 是日较差。为了找出感染性腹泻疾病发病与环境气象因子之间的相关关系,采用经典方差贡献分析技术,现将部分计算结果列于表 2。因子  $X_i$  的方差贡献值  $PV_i$ ,即因子的偏回归平方和,是因子重要性的度量方法之一,常用式(1)计算。

$$PV_i = B_i^2 / C_{ii} / L_{yy} \quad (1)$$

式(1)中, $i=1,2,3, \dots, m$ ,  $B_i^2$  为因子  $X_i$  的偏回归平方和,  $L_{yy}$  为感染性腹泻疾病发病的原方差,  $C_{ii}$  为正规方程系数逆矩阵对角线元素。为便于不同因子相互比较,表 2 中,已将其处理为百分比(表 2 中的方差贡献值、总体复相关系数、统计量  $F$  值均是以预测时效 24

~72 小时参数为例)。

从表 2 数据不难看出:感染性腹泻疾病发病与水汽压  $V_p$  相关密切,单相关系数可达 0.8340(当样本数目超过 1000 时,通过信度  $\alpha=0.01$  检验的相关系数下限阈值  $R_\alpha$ ,仅为 0.081),该因子的方差贡献值可占总贡献的 70.6%。加上方差贡献较大平均风速、气温日较差因子,贡献值可升至 93.4%。其次,所有因子的复相关系数平均可达 0.8591,统计量  $F$  值,可达 298.1。远远超过了通过  $F$  检验所需的下限阈值( $\alpha=0.05, F_\alpha=1.83$ )。

表 2 感染性腹泻疾病发病与气象因子的相关分析

环境气象因子 $X_i$	方差贡献值/%	单相关系数
1. 日平均气温 $T_{\text{mean}}(\text{C})$	0.0	0.7645
2. 气温变化倾向 $\Delta T \text{C}$	0.3	-0.0294
3. 气温较差 $T_{\text{max}} - T_{\text{min}} \text{C}$	4.8	-0.2697
4. 平均相对湿度 $RH(\%)$	1.8	0.5986
5. 最小相对湿度 $R_{\text{min}}(\%)$	3.2	0.5760
6. 平均水汽压 $V_p(\text{hPa})$	70.6	0.8340
7. 日降水量 $P_{\text{rec}}(\text{mm})$	1.3	0.4424
8. 平均风速 $W_{\text{ind}}(\text{m} \cdot \text{s}^{-1})$	18.0	-0.3872
9. 平均海平面气压 $AP(\text{hPa})$	0.0	-0.6642
10. 气压变化倾向 $\Delta H_{\text{pa}}$	0.0	0.0416
总体复相关系数		0.8591
统计量 $F$ 值		298.15

这就证明,感染性腹泻疾病发病与气象因子群之间是少见的高度相关,利用环境气象因子的变化预测北京地区未来感染性腹泻疾病发病趋势,是完全可行的。

通常,感染性腹泻疾病的病菌从适宜的温湿环境中生长繁殖到通过苍蝇等传播媒介污染生冷食物,最终感染人体。人们生活实践的常识,认为连续的高温高湿的气象条件非常易于食物霉变、并会促进病菌的繁殖和延长病菌的生存时间,通过污染的手、食品、水源或苍蝇、蟑螂等间接媒介传播最终经口入消化道使易感者受感染;又因为天气闷热,人们喜欢大量食用生、冷食品使胃肠道防御功能降低导致感染性腹泻的发病流行。感染性腹泻疾病发病的过程反映了水汽压在诸气象因子中贡献最大的特点。

## 2 医疗气象预报建模与评估疾病流行风险水平的思路与方法

感染性腹泻病医疗气象预报与流行风险水平评估采用 5 档分级的方法。不同的级别,对应不同的疾病发病人数及流行风险水平。为此,通过下面两个步骤来完成。

### 2.1 组建逐日发病人次滚动预报的多元回归模式

本文,采用经典回归技术建立预报模型。该模式可写为下式,即

$$Y_i = B_0 + \sum B_j * X_j \quad (2)$$

式(2)中, $Y_i$ 为每百万人口的平均日发病人次的回归拟合值。 $X_j$ 为进入方程的预报因子。 $B_j(j = 1, 2, \dots, m)$ 为相应的回归系数、 $B_0$ 为常数系数,可通过求解正规方程获取。

经典回归模式的正规方程为:

$$\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{X}'\mathbf{Y} \quad (3)$$

式(3)中, $\mathbf{X}'\mathbf{X}$ 为因子协方差矩阵,可以根据因子的方差贡献的大小预选  $n$  个因子组合搭

配建模,或采用逐步回归技术自动筛选。根据我们的试验,后者统计味道过浓,而前者可以综合考虑压、温、湿、风不同量纲因子搭配组合,预报方程稳定,效果更佳。 $\mathbf{X}'\mathbf{Y}$ 为相应气象因子与感染性腹泻疾病的交叉平方和, $\mathbf{B}$ 为回归系数向量。具体建模之前,需从已经搜集的原始样本序列中,连续滑动地截取预报单元( $P+T$ ),作为对应不同预报时效的有效样本备用。前者对应过去  $P$  天的气象因子连续取值区间,后者对应未来  $T$  天平均发病的取值区间。不同建模时效对应的各年最大可能样本数为:  $365 + L_p - (P+T) + 1$ 。闰年  $L_p = 1$ , 否则为 0。我们共做了 6 种时效的建模试验。在此,使用 6 个精选气象预报因子(气温  $X_1$ 、气温较差  $X_3$ 、最小相对湿度  $X_5$ 、水汽压  $X_6$ 、平均风速  $X_8$ 、海平面气压  $X_9$ ),仅给出预报未来 72 小时时效的感染性腹泻疾病的回归预测模式(其他 5 个时效同样用该 6 个预报因子):

$$\begin{aligned} Y_{72} = & -365.7268 + 0.05503434X_1 - \\ & 0.6254867X_3 - 0.6541398X_5 + \\ & 4.133605X_6 - 4.797883X_8 + \\ & 0.3904339X_9 \end{aligned} \quad (4)$$

式(4)中的  $Y_{72}$  为感染性腹泻疾病的未来 72 小时百万人口发病人次的预测值。模式的复相关系数为 0.8677,统计量  $F$  值为 539.66。

### 2.2 疾病预报分级与流行风险水平的评估

同步评估疾病流行风险水平与预报结果的气象分级,应用概率积分的方法进行评估与分级。该方法的要点是:

首先,按照评估疾病流行风险水平(用 4 种医学、气象概率密度分级试验)的实际需要,确定不同的分级概率标准。比如,概率标准定为 5 级,各级所占概率分别为  $P_5$ 、 $P_4$ 、 $P_3$ 、 $P_2$  和  $P_1$ 。注意, $\sum P_i = 1 (i = 5, 4, \dots, 1)$  这将意味着式(2)的预报结论中,将有  $P_i \times 100\%$  的预报结果可以期望降落到第  $i$  级。

其次,如何判断式(2)预报结果  $Y_i$  的级别归属? 如果仍将概率标准定为 5 级,在样本的连续域上,为了分 5 个等级,须有 4 个分点即  $g_4$ 、 $g_3$ 、 $g_2$  和  $g_1$ 。

当  $Y_i > g_4$  预报 5 级。发病人数将显著增多,流行风险水平很高。

当  $Y_i > g_3$  and  $Y_i \leq g_4$  预报 4 级。发病人数较多,流行风险水平较高。

当  $Y_i > g_2$  and  $Y_i \leq g_3$  预报 3 级。发病人数呈现增多趋势,流行风险水平中等

当  $Y_i > g_1$  and  $Y_i \leq g_2$  预报 2 级。发病人数较少,流行风险水平较低。

当  $Y_i < g_1$  预报 1 级。发病人数很少,流行风险水平很低。

分界指标  $g_1$ 、 $g_2$ 、 $g_3$  和  $g_4$  的获取,可以根据指定的概率密度标准  $P_1$ 、 $P_2$ 、 $P_3$ 、 $P_4$ 、

$P_5$ ,对式(2)作概率积分,当曲边梯形的面积,分别达到  $\sum P_i$  时,所记录的 4 个分点位置  $g_i$  ( $i = 1, 2, 3, 4$ ),即为 5 级的分界指标。余可类推,不再赘述。

### 3 结果分析

#### 3.1 所建模型对历史资料的回代模拟

首先,指定一种预报时效及相应的概率分级标准,对所有历史样本实况进行同步分级(含流行风险水平评估,下同),再用同期气象因子值回代式(2),获得所有建模样本的拟合预报分级。现以概率密度标准:0.15 0.20 0.30 0.20 0.15 为例(过去 7 天气象资料,预报未来 6 种时效),拟合结果的比较见表 3。

从表 3 数据可以看出,如果以预报结果

表 3 不同预报时效、不同病种历史样本拟合率的比较

预测时效/小时	样本数	完全正确/%	相差±1级/%	正确率(含±1级)/%	失误差/%
24~48	1069	50.879	43.756	94.635	5.365
24~72	1066	51.763	43.506	95.269	4.731
24~96	1063	51.256	44.837	96.093	3.907
24~120	1060	50.093	46.362	96.455	3.545
24~144	1057	51.169	46.305	94.474	2.526
24~168	1054	51.220	46.623	97.842	2.158

与实况相差±1级以内算作正确的话,不同时效的拟合正确率,平均可达 95%以上。预报级别与实况完全一致的,约占 51%左右。

#### 3.2 不同概率密度标准对拟合准确率的可能影响

在 5 级分档概率密度标准中,如果适当降低两端概率密度  $P_1$ 、 $P_5$  的取值,即调整到医学常用的概率密度分布函数:0.03594、0.23831、0.4515、0.23831、0.03594。注意,需保证  $\sum P_i = 1$ 。仍使用过去 7 天气象资料,

预报未来 2~7 天的模型,则拟合准确率有如下变化,详见表 4。

表 4 中,是针对两端概率适当下调,总的拟合正确率平均提高了大约 1~2 个百分点;级别完全符合的百分率,平均可提高 14 个百分点;相差±1级的百分率,平均降低了 11 个百分点。经过这样的调整,拟合准确率有明显改善,但预报产品结果中,出现在 1 级或 5 级的心理期望概率,必然有所降低,是符合感染性腹泻疾病的流行风险很大的水平为小概率事件现象。

表 4 调整概率密度标准后的感染性腹泻疾病历史样本拟合率

预测时效/小时	样本数	完全正确/%	相差±1级/%	正确率(含±1级)/%	失误差/%
24~48	1069	62.905	35.245	98.150	1.850
24~72	1066	63.265	35.158	98.423	1.577
24~96	1063	64.744	33.674	98.419	1.581
24~120	1060	65.299	33.489	98.787	1.213
24~144	1057	66.137	32.834	98.971	1.029
24~168	1054	66.417	32.083	98.499	1.501

#### 4 服务应用检验

应用感染性腹泻疾病预测模式,预测出

未来 6 种时效的逐日的感染性腹泻疾病发病等级与北京市海淀区 2007 年逐日监测病例比较,统计出各时效的感染性腹泻疾病预报准确率,其效果是理想的,见表 5。

表 5 北京市海淀区 2007 年感染性腹泻病预测试验准确率

预报时效/小时	样本数	完全正确/%	误差±1级/%	误差≤1级/%	误差>1级/%
24~48	356	81.18	18.54	99.72	0.28
24~72	355	62.25	36.06	98.31	1.69
24~96	354	62.15	36.16	98.31	1.70
24~120	353	61.19	37.11	98.30	1.70
24~144	352	61.08	36.65	97.73	2.27
24~168	351	58.12	39.60	97.72	2.28

#### 5 结语

本研究方法同步应用软件通用性较强,只要搜集到当地较大样本发病资料,即可运行建模,并试验发布预报。当取得更多资料时,还可对原有模型参数优化。目前,感染性腹泻病疾病医疗气象预报软件,经过气象科技服务业务的试验应用,感觉操作逻辑流畅,产品丰富图文并茂,还可以方便查询各种相关信息,预测效果较为理想。在奥运赛期,与北京市海淀区疾病预防控制中心合作,为奥运组委会提供较准确的感染性腹泻发病趋势预测产品,也可以为首都的市民提供预防

感染性腹泻病的科学信息,提高百姓的身体健康水平。

#### 参考文献

- [1] 张德山,王玉玲,谢庄. 天气变化影响人体健康的信息与预报[J]. 气候与环境研究,1999,4(1):32-34.
- [2] 张庆阳,琚建华,王卫丹. 气候变暖对人类健康的影响[J]. 气象,2007,35(2):245-248.
- [3] 张德山,邵明轩,穆启占,等. 密云水库流域暴雨的短期预报方法研究[J]. 气象,2006,32(11):61-64.
- [4] 屠其璞,王俊德,丁裕国,等. 气象应用概率统计学[M]. 北京:气象出版社,1984.
- [5] 宛公展,王宝成,弥明祥. PC-1500 袖珍计算机 BASIC 语言应用程序集[M]. 北京:气象出版社,1988:132.