

侯宜广, 赵瑾, 董玉昆, 等. 利用 Kappa 统计值研究多种空气质量预报方法的一致性[J]. 气象, 2011, 37(2): 232-236.

利用 Kappa 统计值研究多种 空气质量预报方法的一致性

侯宜广 赵瑾 董玉昆 孙建印

江苏省徐州市气象局, 徐州 221002

提 要: 采用基于权重 Kappa 统计值的方法, 在剔除了由于偶然性和随机性造成的一致性的基础上, 对三种常用的空气质量预报方法的预测结果的一致性进行了衡量。结果表明, “数值模式预报”与“综合经验预报”的预测结果为“中等的一致性”, 说明其预测结果的一致性并不是由于偶然性造成的。衡量结果是清晰的, 避免了来自主观评价的差异。从而, 有利于提高对不同模型预测结果的差异性的认识, 对进一步提高空气质量预报的准确度有着重要意义。

关键词: 空气质量预报, 权重 Kappa 统计值, 一致性

Study on the Consistency of Air Quality Forecasting Methods Using Kappa Statistic

HOU Yiguang ZHAO Jin DONG Yukun SUN Jianyin

Xuzhou Meteorological Office of Jiangsu Province, Xuzhou 221002

Abstract: Using Kappa value and taking out the impact of chance and random this paper measures the inter-agreement of predictive results of three different air quality forecasting methods. The result shows that, the predicting results of both numerical model forecasting and synthesis experience forecasting are of medium identities, and they are not casual. The results measured are clear, and the differences from subjective assessment can be avoided. Furthermore, this research would promote the understanding of different forecasting models and be important to the improvement of forecasting accuracy.

Key words: air quality forecasting, weighted Kappa value, inter-agreement

引 言

现代工业及交通运输业等的快速发展, 使得城市空气污染日益严重, 越来越影响人们的健康和生活^[1-3]。环境恶化也使人们越来越认识到保护人类赖以生存的大气环境的重要性与紧迫感^[3]。空气质量预报的发布在社会公众中已经引起了较大反响, 对宣传环保和保护环境也起到了积极的推动作用^[1]。空气质量预报是复杂的系统工程, 也是环境科学研究的热点和难点所在。

许多学者和从业人员已经对空气质量预报的方

法进行了有益的探讨。房小怡等介绍了一种城市空气质量数值预报模式系统, 该系统由气象模式(中尺度气象模式 MM5 及城市大气边界层模式 CBLM)和大气污染物输送扩散化学模式组成^[4]。尤焕苓等通过分析北京地区最大混合层厚度与空气质量的关系提出污染物在铅直方向稀释、扩散的最大范围, 是污染预测的指标之一, 也是空气污染潜势预报的重要参数^[5]。郑选军等以可吸入颗粒物级别预报为例, 提出了回归诊断在城市空气质量预报中的应用方法, 以减少回归系数估计的误差, 提高模型预报的精度^[6]。阴俊等指出分类统计预报模型可以取得较高的实际业务预报精度, 较全样本统计预报能力有

所提高^[7]。吴振玲等使用短期气候趋势预报原理与短期空气质量预报相结合的方法,对冬季采暖期的空气质量进行预测,不仅借鉴了传统的短期天气气候预测方法和经验,而且还充分地发挥了目前短期空气质量统计预报方法的作用^[8]。许建明等利用不同季节 CMAQ 模式产品与观测实况资料,建立了 CMAQ-MOS 区域空气质量统计修正预报模型,并采用检验方法评估 CMAQ-MOS 方案预报能力,提出采用 CMAQ-MOS 统计修正模型统计-动力相结合的空气质量预报新途径^[9]。徐晓峰等介绍了空气污染预报业务流程,详细介绍了根据这个业务流程所开发的北京市气象局空气污染预报业务软件系统^[10]。王自发等综述了中国科学院大气物理研究所自主开发的嵌套网格空气质量预报模式系统的历史发展与应用情况^[11]。

在认识到各种空气质量预报的方法都有各自的适用性和不足以后,相关研究开始针对不同的空气质量预报方法进行了比较和分析。朱玉强对常用的空气质量预报方法:数值模式预报、统计预报和综合经验预报进行了比较和分析^[12]。孙峰介绍了北京市空气质量动态统计预报系统的构成,以及所采用的线性回归模型 LRM,分类判别树 CART 模型, CART 与 LRM 结合的模型,动态统计预报模型 DSM,多点预报模型 MPDSM 5 种预报模型,分析了不同预报模型的特点和性能^[13]。周秀杰等提出了空气污染指数 BP 网络预报模型,结果表明 BP 方法预报模型的预报准确度明显高于通常使用的逐步回归方法,特别是对骤升骤降趋势也能得到准确度较高的预测结果^[3]。吕蓓红等选用统计预报模式(多元线性回归)和数值预报模式(CAPPS)两种模式同步应用,提高了安阳市在空气质量预报方面准确率^[14]。李金义等建立了简单空气质量模式系统(SAQMS),并与中科院大气物理研究所研制的 HRCM 模式系统的预测结果进行比较^[15]。高安春等利用逐步回归筛选的因子及统计模型研究中的有关数据,指出使用卡尔曼滤波方法制作空气污染物浓度预报可以取得比较满意的效果^[16]。姜有山等利用高空大气探测资料和有限区细网格模式的数值预报产品,分别建立统计预报和数值预报等客观预报方法,指出动态分权法能够有效地提高空气质量的预报准确率^[1]。喻雨知等采用统计和对比方法,分析了长沙市空气质量现状,介绍了天气形势相似及动态逐步回归两种空气质量统计预报方法,并对

其一年多的运行结果进行了检验和对比^[2]。

然而,在不同预测模型下的预测结果所达成的一致性(即预测结果相同),也有可能是由于偶然性和随机性所造成的,现有的文献中缺乏针对解决此类问题的相关研究。换句话说,如何在比较不同预测模型的预测结果的时候,剔除偶然性和随机性的影响,在空气质量预报领域仍然是个空白。本文的研究目的就是基于 Kappa 统计值的方法,在剔除由于偶然性和随机性造成的一致性的基础上,对三种常用的空气质量预报方法(即数值模式预报、统计预报和综合经验预报)的预测结果的一致性进行衡量,从而提高对不同模型预测结果的差异性的认识。此项研究对于更加深刻地认识各种预测模型及其预测结果、对进一步提高空气质量预报的准确度有着重要意义。

1 研究方法及理论基础

1.1 Kappa 统计值概述

Kappa 统计值是被广泛采用的、用来衡量去除了偶然性以后的内部一致性的统计量^[17-20]。这种方法在临床诊断、分类等领域得到了非常广泛的应用^[19, 21]。这种方法考虑了由于偶然性达成一致的可能性,通过首先从表面一致性里面减去由于偶然性导致一致性的可能性,然后再除以非偶然性导致的一致性,从而计算出相应的 Kappa 统计值^[22-23]。采用的预测结果的分类为 4 个:优良、轻度污染、中度污染和重度污染^[3],则每个预测模型的结果都有 4 种可能。显然,距离远的两个选项的差异性要比距离近的两个选项的差异性要大,为了反映这种差异性,就要给每一个情况赋以不同的权重,权重 Kappa 统计值就可以解决这种情况。样本数可根据 Kappa 精度要求来确定,通常需要 20~30 个样本量^[24]。样本量不需要太大,当样本量很大,且一致率较高时,即使计算结果有统计学意义,其实际意义也不大。

1.2 权重 Kappa 统计值

1.2.1 衡量两种预测模型的一致性

由于理论上的一些原因, Fleiss 建议为每对比较选项按照以下的公式赋以权重^[22]:

$$\omega_{ij} = 1 - \frac{(i-j)^2}{(k-1)^2} \quad (1)$$

这也是被广泛采用的赋权重方法之一,式(1)中

i 表示行数, j 表示列数, k 表示选项个数, 下同。表面一致性通过以下公式计算:

$$p_o(\omega) = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^k \omega_{ij} S_{ij} \quad (2)$$

式(2)中 N 表示样本总数, ω_{ij} 表示所赋权重值, S_{ij} 表示预测结果分组汇总的二维分布。

由偶然性导致的一致性计算公式如下:

$$p_e(\omega) = \frac{1}{N^2} \sum_{i=1}^k \sum_{j=1}^k \omega_{ij} S_i S_j \quad (3)$$

式(3)中 N 表示样本总数, ω_{ij} 表示所赋权重值, S_i 表示分组中一种预报方法的 i 分类之和, S_j 表示分组中另一预报方法的 j 分类之和。

则权重 Kappa 值为:

$$k(\omega) = \frac{p_o(\omega) - p_e(\omega)}{1 - p_e(\omega)} \quad (4)$$

对 Kappa 统计值的解释一直是研究人员关心的热点问题。就 Kappa 值的意义而言, 它的变化范围是从 -1 到 +1。在小于 0 的情况下, 表明结果是由于偶然性造成的, 等于 1 则表示“完美的一致性”。表 1 表示了对 Kappa 统计值进行解释的常用的指导性标准^[18-19]。

表 1 对 Kappa 统计值进行解释的指导性标准

Table 1 Guidance standard of elucidation to Kappa statistic

| Kappa 统计值 | 并不是由于偶然性造成的一致性的支持强度 |
|-----------|------------------------|
| <0 | 很差 (Poor) |
| 0~0.20 | 微弱的 (Slight) |
| 0.21~0.40 | 清楚的 (Fair) |
| 0.41~0.60 | 中等的 (Moderate) |
| 0.61~0.80 | 本质的 (Substantial) |
| 0.81~1.00 | 几乎完美的 (Almost perfect) |

1.2.2 衡量多种预测模型的一致性

当涉及到多个模型、多个预测结果时, 最常见的做法是遍取所有可能的不同两个预测模型之间的 Kappa 统计值, 然后取所有 Kappa 统计值的平均值。用公式表示则是:

$$k_{ave}(\omega) = \frac{1}{Q} \sum_{q=1}^Q k_q(\omega) \quad (5)$$

Fleiss 建立了另外一种一次性计算该种情况下的 Kappa 统计值^[22]的模型, 但并没有得到广泛的应用。其最大的问题在于对结果的解释上, 因为不一致性会随着选项的增多以及评价人数的增多而减小, 按照这种方法计算出来的 Kappa 值相对较小, 不太适合去和一般性意义下的 Kappa 统计值的指导标准进行比较(参见表 1)。这也是样本量不要

求很大的主要原因。

2 案例分析及计算结果

2.1 案例背景

朱玉强指出, 目前应用于我国各个城市空气质量预报业务的预报方法主要有三种: 数值模式预报、统计预报和综合经验预报^[12]。我们选取了 2008 年 6 月 1 日至 20 日由徐州市环境监测中心站与徐州市气象台利用这三种方法联合制作的徐州市空气质量预报结果, 预报结果汇总如表 2 所示。

表 2 三种模型的预测结果示例

Table 2 Examples of predicting results of three models

| 时间点 | 数值模式 | 统计预报 | 综合经验 | 时间点 | 数值模式 | 统计预报 | 综合经验 |
|-----|------|------|------|-----|------|------|------|
| 1 | II | II | I | 11 | III | II | II |
| 2 | II | I | I | 12 | II | II | II |
| 3 | II | II | I | 13 | II | II | II |
| 4 | I | II | I | 14 | II | I | II |
| 5 | I | II | I | 15 | II | I | II |
| 6 | II | II | I | 16 | III | II | II |
| 7 | II | II | II | 17 | II | I | I |
| 8 | I | I | I | 18 | II | I | I |
| 9 | I | I | I | 19 | II | I | I |
| 10 | I | II | I | 20 | I | II | I |

注: I: 优良 (API: 0~100); II: 轻度污染 (API: 100~200); III: 中度污染 (API: 200~300); IV: 重度污染 (API: >300)^[3]。其中, “数值模式”是采用“城市空气污染数值预报系统 CAPPS”^[25-26]本地化制作预报, “统计预报”是采用回归相关法制作预报, “综合经验”是利用业务人员长期经验总结作出预报

2.2 权重的赋值

按照公式(1)对不同对比选项组赋以的权重如表 3 所示。

表 3 权重 Kappa 统计值中赋的权重

Table 3 Weights given in Kappa statistic

| | I | II | III | IV |
|-----|-------|-------|-------|-------|
| I | 1.000 | 0.889 | 0.556 | 0.000 |
| II | 0.889 | 1.000 | 0.889 | 0.556 |
| III | 0.556 | 0.889 | 1.000 | 0.889 |
| IV | 0.000 | 0.556 | 0.889 | 1.000 |

注: I: 优良 (API: 0~100); II: 轻度污染 (API: 100~200); III: 中度污染 (API: 200~300); IV: 重度污染 (API: >300)^[3]

2.3 权重 Kappa 值的计算

表 4 是根据表 2 中三种预报方法的预测结果汇总统计出来的交叉表。

表4 三种预报方法的预测结果汇总交叉表
Table 4 Crossing summary in the predicting results of three forecasting methods

| | | 数值模式预报 | | | | |
|--------|-------|----------|----|----|-----|----|
| | | S_{ij} | I | II | III | IV |
| 统计预报 | I | 3 | 6 | 0 | 0 | 9 |
| | II | 3 | 6 | 2 | 0 | 11 |
| | III | 0 | 0 | 0 | 0 | 0 |
| | IV | 0 | 0 | 0 | 0 | 0 |
| | S_j | 6 | 12 | 2 | 0 | |
| 综合经验预报 | I | 6 | 7 | 0 | 0 | 13 |
| | II | 0 | 5 | 2 | 0 | 7 |
| | III | 0 | 0 | 0 | 0 | 0 |
| | IV | 0 | 0 | 0 | 0 | 0 |
| | S_j | 6 | 12 | 2 | 0 | |
| | | 统计预报 | | | | |
| | | S_{ij} | I | II | III | IV |
| 综合经验预报 | I | 6 | 7 | 0 | 0 | 13 |
| | II | 2 | 5 | 0 | 0 | 7 |
| | III | 0 | 0 | 0 | 0 | 0 |
| | IV | 0 | 0 | 0 | 0 | 0 |
| | S_j | 8 | 12 | 0 | 0 | |

注: I: 优良(API: 0~100); II: 轻度污染(API: 100~200); III: 中度污染(API: 200~300); IV: 重度污染(API: >300)^[3]

根据表3中的 ω_{ij} 权重值和表4中的 S_{ij} , S_i , S_j 的统计值,按照式(2)、式(3)和式(4),三种预报方法之间的表面一致性、由于偶然性造成的一致性、Kappa统计值的计算如表5所示。

表5 三种预报方法之间的一致性衡量计算结果
Table 5 Comparison of measured and calculated identities by three forecasting methods

| | “数值模式 预报”和 “统计预报” | “数值模式 预报”和“综 合经验预报” | “统计预报” 和“综合经 验预报” |
|------------------|-------------------------|---------------------------|-------------------------|
| 表面一致性 | 0.939 | 0.950 | 0.950 |
| 偶然一致性 | 0.926 | 0.912 | 0.941 |
| Kappa统计值 (含义) | 0.179 (微弱的一致性) | 0.430 (中等的一致性) | 0.151 (微弱的一致性) |

从表5中的计算结果可以看出,“数值模式预报”和“统计预报”的表面一致性高达0.939,而剔除偶然性造成的一致性后,Kappa统计值仅仅表示“微弱的一致性”。同理,“数值模式预报”和“综合经验预报”为“中等的一致性”,说明其预测结果的一致性并不是由于偶然性造成的。“统计预报”和“综合经验预报”之间也仅仅是“微弱的一致性”。

3 结论

针对现有研究的不足,本文基于权重 Kappa 统

计值的方法,在剔除了由于偶然性和随机性造成的一致性的基础上,对三种常用的空气质量预报方法的预测结果的一致性进行了衡量。结果表明,“数值模式预报”和“统计预报”的表面一致性高达0.939,而剔除偶然性造成的一致性后,Kappa统计值仅仅表示“微弱的一致性”。同理,“数值模式预报”和“综合经验预报”为“中等的一致性”,说明其预测结果的一致性并不是由于偶然性造成的。“统计预报”和“综合经验预报”之间也仅仅是“微弱的一致性”。可以看出,这种统计学方法可同时排除偶然一致性和系统性误差的干扰,要比一般的相关系数更好^[24]。采用基于权重 Kappa 统计值的方法来衡量这三种预报方法的结果是清晰的,也是可信的,避免了来自主观评价的差异。从而,有利于提高对不同模型预测结果的差异性的认识,对进一步提高空气质量预报的准确度有着重要意义。

参考文献

- [1] 姜有山,陈飞,班欣,等. 连云港市城市空气质量预报方法研究[J]. 气象科学,2007,27(2):220-225.
- [2] 喻雨知,王体健,肖波,等. 长沙市两种空气质量预报方法检验对比[J]. 长江流域资源与环境,2007,16(4):509-513.
- [3] 周秀杰,苏小红,袁美英. 基于BP网络的空气污染指数预报研究[J]. 哈尔滨工业大学学报,2004,36(5):582-585.
- [4] 房小怡,蒋维楣,吴润. 城市空气质量数值预报模式系统及其应用[J]. 环境科学学报,2004,24(1):111-115.
- [5] 尤焕苓,刘伟东,谭江瑞. 北京地区平均最大混合层厚度的时间变化特征[J]. 气象,2010,36(5):51-55.
- [6] 郑选军,王国强. 回归诊断在城市空气质量预报中的应用研究[J]. 气象,2004,30(9):9-13.
- [7] 阴俊,谈建国. 上海城市空气质量预报分类统计模型[J]. 气象科技,2004,32(6):410-413.
- [8] 吴振玲,谢以扬,周惠,等. 2003年冬季空气质量趋势预测方法[J]. 气象,2005,31(10):47-50.
- [9] 许建明,徐祥德,刘煜,等. CMAQ-MOS区域空气质量统计修正模型预报途径研究[J]. 中国科学D辑(地球科学),2005,35(增刊1):131-144.
- [10] 徐晓峰,赵习方,张小玲,等. 北京空气污染预报业务系统[J]. 气象,2005,31(1):63-66.
- [11] 王自发,谢付莹,王喜全,等. 嵌套网格空气质量预报模式系统的发展与应用[J]. 大气科学,2006,30(5):778-790.
- [12] 朱玉强. 几种空气质量预报方法的预报效果对比分析[J]. 气象,2004,30(10):30-33.
- [13] 孙峰. 北京市空气质量动态统计预报系统[J]. 环境科学研究,2004,17(1):70-73.

- [14] 吕蓓红,曹三忠,张勇. 环境空气质量预报准确率研究[J]. 信阳师范学院学报(自然科学版),2006,19(4):499-503.
- [15] 李金义,陈万隆,冯怀莹,等. 沈阳市区 2 种空气质量数值模式预报结果的对比分析[J]. 安徽农业科学,2006,34(11):2614-2617.
- [16] 高安春,申培鲁. 利用 MM5 模式输出产品制作空气质量预报方法探讨[J]. 气象科学,2007,27(1):57-62.
- [17] Viera A J, Garrett J M. Understanding interobserver agreement: The Kappa statistic [J]. Family Medicine, 2005, 37(5):360-363.
- [18] Vries H D, Elliott M N, Kanouse D E et al. Using pooled Kappa to summarize interrater agreement across many items [J]. Field Methods, 2008, 20(3): 272-282.
- [19] Gambatese J A, Behm M, Rajendran S. Design's role in construction accident causality and prevention: Perspectives from an expert panel [J]. Safety Science, 2008, 46: 675-691.
- [20] 吴伟巍. 施工现场安全危险源实时监控与安全风险预测方法研究[D]. 东南大学博士学位论文,2009.
- [21] Kundel H L, Polansky M. Measurement of observer agreement [J]. Radiology, 2003, 228: 303-308.
- [22] Fleiss J L, Levin B, Paik M C. Statistical Methods for Rates and Proportions (Third Edition) [M]. John Wiley & Sons, Inc. 2003.
- [23] Feinstein A, Cicchetti D. High agreement but low Kappa. I. The problem of two paradoxes[J]. J Clin Epidemiol. 1990, 43:543-549.
- [24] 吴毅,胡永善,范文可,等. 功能评定量表信度和效度的研究[J]. 中国康复医学杂志,2004,19(3):230-231.
- [25] 朱蓉. 城市空气污染数值预报系统 CAPPs 及其应用[J]. 应用气象学报,2001,12(3):267-278.
- [26] 刘宁微,马雁军. 区域空气质量数值预报系统 CAPPs3 在辽宁的应用[J]. 气象,2009,35(10):84-89.