

# 基于气象资料的林芝地区空气质量动态 预报方法研究

陈宫燕<sup>1</sup> 德吉白玛<sup>1</sup> 红梅<sup>1</sup> 旺杰<sup>1</sup> 旦增<sup>1</sup> 宋燕<sup>2</sup>

(1 林芝地区气象局, 林芝 860000; 2 中国气象局气象干部培训学院, 北京 100081)

**摘要:** 空气污染状况不仅与污染源排放有关, 也与气温、降水等气象要素存在密切联系。建立区域气象资料与空气质量的关系, 对研究空气质量及其变化具有重要的研究意义。由于空气污染与气象条件之间并非线性关系, 将前20d的气象观测资料和实际的大气污染物浓度作为输入参数, 动态建立每天的多元线性回归方程, 代入预报气象数据求取SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>的预测值和空气污染指数(API)值。初步试报表明, 以20d为时间窗进行滑动预报, 充分考虑了空气污染与气象条件之间的复杂动态关系, 克服了传统静态空气质量预报方法的缺点, 能更高精度地预报林芝地区的空气质量, 具有一定的应用和推广价值。

**关键词:** 空气质量动态预报, 多元线性回归模型, 污染物浓度

DOI: 10.3969/j.issn.2095-1973.2013.06.008

## Air Quality Dynamic Prediction Method Based on Nyingchi Meteorological Data

Chen Gongyan<sup>1</sup>, De Ji Baima<sup>1</sup>, Hong Mei<sup>1</sup>, Wang Jie<sup>1</sup>, Dan Zeng<sup>1</sup>, Song Yan<sup>2</sup>

(1 Meteorological Bureau of Nyingchi, Nyingchi 860000 2 CMA Training Centre, China Meteorological Administration, Beijing 100081)

**Abstract:** Air pollution is not only associated with emissions of pollutant sources, but also has close links with meteorological factors such as temperature and precipitation. Establishment of the relationship between regional meteorological data and air quality is of great importance of the study of air quality and its changes. Because of the non-linear relationship between air pollution and meteorological conditions, the meteorological data and the actual concentration of air pollutants 20 days before as an input parameter dynamically created the daily multiple linear regression equation, then substituted the forecast meteorological data with the obtained equation to solve the predictive values and air pollution index (API) values of SO<sub>2</sub>, NO<sub>2</sub>, and PM<sub>10</sub>. The preliminary test prediction demonstrated that the sliding forecast method of 20 days as a time window could more accurately forecast the air quality of the Nyingchi and had certain application and promotion values. By using this method, meteorologists took adequately into account the complex and dynamic relationship between air pollution and meteorological conditions, thus overcoming the shortcomings of the traditional static air quality forecasting method.

**Keywords:** air quality prediction, multiple linear regression model, pollutant concentration

目前我国空气质量的预报方法有数理统计、数值模式和综合经验3类方法<sup>[1]</sup>。气象条件对排入到大气中的污染物有明显的的作用, 二者之间存在一定的相关性。数理统计方法即是利用这种相关性, 通过分析自变量与因变量之间的统计规律, 寻找它们之间的关系。由于该方法仅对观测数据进行相关性分析就可得出自变量与因变量的近似函数关系, 因此应用起来比较简便, 也是空气质量预报的主要方法。

数理统计方法目前的主要思路是根据长时间序列

的气象资料, 建立静态的多元线性回归方程<sup>[2]</sup>。由于空气污染是一个复杂的过程, 与气象条件之间的关系并非线性过程, 静态的拟合函数并不能得到高精度的预报结果。因此, 需要建立动态空气质量预报方法, 以提高空气质量预报精度。

### 1 研究区域空气质量概况

林芝地区地处西藏东南部, 气候宜人、环境优美、物产丰富, 素有“西藏江南”之美称, 是全球生态环境最好的地区之一, 被称为中国乃至东半球气候的“启动器”和“调节区”<sup>[3]</sup>。该区域常年受南亚高压和孟加拉湾暖湿气流的影响, 形成了复杂多样的独特气候, 正是这些独特的气候资源造就了西藏林芝的

收稿日期: 2013年1月5日; 修回日期: 2013年6月8日

第一作者: 陈宫燕(1984—),

Email: chengongyan\_35@163.com

碧水蓝天。然而，它的生态系统又是世界上最脆弱的地区之一<sup>[4]</sup>，需要得到人类的呵护。

林芝地区环境监测站自2009年9月开始开展污染物浓度监测。根据国家环保局规定，空气质量主要以SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>三种污染物浓度来衡量。从图1可见，影响林芝地区八一镇空气的主要是PM<sub>10</sub>，PM<sub>10</sub>的值在冬春季最高，夏季最低，呈两高一低的变化，两个峰值区出现在3月和11月，分别为0.035和0.032，低值区出现在7月，为0.014；而SO<sub>2</sub>、NO<sub>2</sub>的值变化幅度不大，SO<sub>2</sub>最高值出现在9月，值为0.01，之后呈下降趋势；NO<sub>2</sub>从10月开始呈上升趋势，最高值出现在11月，值为0.011。

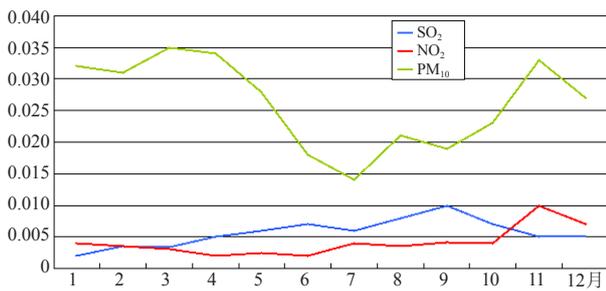


图1 林芝地区八一镇2011年SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>浓度值月平均曲线(单位: mg/m<sup>3</sup>)

## 2 资料和方法介绍

本文采用林芝地区环境监测站提供的2011年1月—2012年5月的监测项目(SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>)的浓度值以及同时段的地面测站观测资料和高空Tlogp资料来做相关分析、回归模型的建立以及对模型的精度检验。

首先，对2011年全年的地面测站观测资料做因子筛选，把选出来的地面测站气象因子和高空T-logP资料进行相关性分析。然后，对相关系数做显著性检验，选出最终相关系数显著的因子来建立回归方程。最后，利用2011年5月1日—2012年4月30日这一时段的数据对建立的回归模型进行误差分析以及动态预报方程的精度评价。

## 3 基于滑动时间窗的空气质量动态预报

### 3.1 多元线性回归模型

在回归分析中，如果有两个或两个以上的自变量，就称为多元回归。事实上，一种现象常常是与多个因素相联系的，需要由多个自变量的最优组合共同来预测或估计因变量，比只用一个自变量进行预测或估计更有效，更符合实际。因此多元线性回归比一元线性回归的实用意义更大。多个(至少两个)要素之间存在着相关影响、相互关联的情况。因此，多元线性回归模型更带有普遍性的意义。

假设某一因变量 $y$ 受 $k$ 个自变量 $x_1, x_2, \dots, x_k$ 的影响，其 $n$ 组观测值为 $y_i, x_{1i}, x_{2i}, \dots, x_{ki}, i=1, 2, \dots, n$ 。那么，多元回归模型的结构形式为

$$y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \dots + a_kx_{ki} + w \quad (1)$$

式中， $a_0, a_1, \dots, a_k$ 为待定参数， $w$ 为随机变量。如果 $b_0, b_1, \dots, b_k$ 分别为 $a_0, a_1, a_2, \dots, a_k$ 的拟合值，则回归方程为

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (2)$$

式中， $b_0$ 为常数， $b_1, b_2, \dots, b_k$ 称为偏回归系数。偏回归系数 $b_i (i=1, 2, \dots, k)$ 的意义是，当其他自变量 $x_j (j \neq i)$ 都固定时，自变量 $x_i$ 每变化一个单位而使因变量 $y$ 平均改变的数值。若引入矩阵，通过计算，正规方程组式可以进一步写成矩阵形式：

$$Ab = B \quad (3)$$

求解式(3)可得：

$$b = A^{-1}B = (X^T X)^{-1} X^T Y \quad (4)$$

### 3.2 滑动时间窗的空气质量动态预报思路

污染物浓度除了受排放量的变化影响外，另一个主要影响因素就是气象要素的变化。有利的气象条件对排入到大气中的污染物有明显的稀释作用，使污染物浓度降低<sup>[5]</sup>。

设预报的大气污染物浓度为 $Y_j$ ，参与预报的气象因子为 $F_i$ ，可建立如下回归方程：

$$Y_j = a_{0j} + \sum_{i=1}^{n_j} a_{ij} F_i \quad (1 \leq j \leq m; 1 \leq i \leq n_j) \quad (5)$$

式中， $a_{ij}$ 为回归系数， $m$ 为污染物数量， $n_j$ 为第 $j$ 个污染物对应的预报因子数量。

将气象观测资料和实际的污染物浓度数据代入公式中，若采样数据数量满足要求，利用式(3)和式(4)可求得回归系数。将气象预报资料和回归系数代入式(5)，得到大气污染物的预测浓度。

传统的预报方法使用长时间序列的样点数据求解各大气污染物浓度的回归方程，并利用该方程完成长时间序列的预报。这种采用相对固定公式预报大气污染物浓度的算法，称为静态预报方法。

由于空气污染的复杂性，气象条件与空气质量之间并非简单的线性关系，静态预报方法随着时间的推移预报误差不断增大。在数学中，常使用微分思想来解决复杂的非线性问题。同理，将气象条件与空气质量放置于时间坐标轴中，对于每天的空气质量预报，采用该天之前的较短时间段(本文以20d为时间窗)的气象观测资料和实际的污染物浓度动态建立污染预报模型。虽然在长时间序列中，气象条件与空气质量非线性关系，但在短时期内可近似看成线性变化。这种以较短时间的线性回归模型模拟长时间序列的气象条件与空气质量非线性关系的方法，称为滑动时间窗

的动态预报方法。

### 3.3 大气污染物主导气象因子的确定

由于气象因子众多，既有气象观测资料，又有数值预报产品，将所有气象因子作为自变量建立与大气污染物浓度的线性关系相对困难，需要选择影响相对较大的部分气象因子进行空气质量预报。而且，不同污染物具有不同的分布特征和扩散规律，其主导气象因子也不一样，因此建立方程前需要为各污染物确定主导气象因子。

采用林芝地区环境监测站提供的2011年1月—2012年5月的监测项目（SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>）的浓度值以及同时段的常规地面测站观测资料和高空T-logp资料做相关分析（表1），总共365个样本。然后，根据表1的相关检验结果，选出相关性非常显著的气象因子作为预报气象因子（表2）。

## 4 误差分析

### 4.1 动态空气污染预报的误差分析

选取2011年5月1日—2012年4月30日的污染物浓度、

地面气象因子、高空T-logp资料，使用动态方程计算出SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>的预报值与实况值进行对比（图2）。可以发现，SO<sub>2</sub>、NO<sub>2</sub>、PM<sub>10</sub>浓度的平均误差分别为0.0009、0.0022、0.0087mg/m<sup>3</sup>，满足空气质量预报要求。从这3种污染物浓度的误差分析可以看出，变化幅度最小的是SO<sub>2</sub>，最大的是PM<sub>10</sub>，说明气象因子以及高空各层的气象资料对PM<sub>10</sub>的影响最大。

### 4.2 与静态空气质量预报方法对比

将2011年5月1日—2012年4月30日的气象数据和污染物观测值建立静态多元回归方程：

$$Y_{11(\text{SO}_2)} = -0.00333 + 0.000816\bar{T} + 0.000163T_{\text{max}} - 0.00259T_{\text{min}} + 0.001822T_{700} + 0.0000964T_{500} - 0.00036T_{400} + 0.000102T_{d700} + 0.000386T_{d500} \quad (6)$$

$$Y_{21(\text{NO}_2)} = -0.88348 - 0.00084\bar{T} + 0.00226T_{\text{min}} + 0.001273\bar{P}_{24} + 0.010024N - 0.01118N_1 - 0.0037T_{700} + 0.000533T_{d700} + 0.001882F_{400} \quad (7)$$

表1 不同污染物各气象因子的相关系数及其检验（采用t检验，粗体为通过0.01显著性水平）

气象因子	SO <sub>2</sub>		NO <sub>2</sub>		PM <sub>10</sub>	
	相关系数	相关系数检验 (t)	相关系数	相关系数检验 (t)	相关系数	相关系数检验 (t)
日平均气温 ( $\bar{T}$ )	0.66	<b>16.738</b>	-0.28	<b>5.557</b>	-0.45	<b>9.601</b>
日最高气温 ( $T_{\text{max}}$ )	0.70	<b>18.675</b>	-0.13	2.498	-0.27	5.343
日最低气温 ( $T_{\text{min}}$ )	0.56	<b>12.878</b>	-0.32	<b>6.435</b>	-0.54	<b>12.224</b>
日降水量	0.05	0.954	-0.02	0.381	-0.30	5.992
日平均气压 ( $\bar{P}_{24}$ )	0.40	8.315	0.30	<b>5.992</b>	0.07	1.337
日相对湿度 (U)	0.36	7.352	-0.14	2.694	-0.58	<b>13.565</b>
日照时数	0.11	2.109	0.13	2.498	0.32	6.435
风速日平均	-0.25	4.919	-0.20	3.889	0.22	4.297
08时能见度	0	0	0.08	1.529	0.24	4.710
08时总云量 (N)	-0.02	0.381	-0.40	<b>8.315</b>	-0.45	<b>9.601</b>
08时低云量 (N <sub>1</sub> )	-0.01	0.191	-0.38	<b>7.827</b>	-0.42	8.817
700hPa温度 ( $T_{700}$ )	0.60	<b>13.930</b>	-0.24	<b>4.592</b>	-0.46	<b>9.623</b>
500hPa温度 ( $T_{500}$ )	0.59	<b>13.570</b>	-0.07	1.303	-0.42	<b>8.596</b>
400hPa温度 ( $T_{400}$ )	0.59	<b>13.570</b>	-0.09	1.678	-0.46	<b>9.623</b>
700hPa露点 ( $T_{d700}$ )	0.57	<b>12.890</b>	-0.23	<b>4.390</b>	-0.45	<b>9.360</b>
500hPa露点 ( $T_{d500}$ )	0.52	<b>11.310</b>	-0.22	4.189	-0.39	7.867
400hPa露点 ( $T_{d400}$ )	0.40	8.106	-0.14	2.626	-0.38	7.631
700hPa风向	0.01	0.186	0.08	1.491	0.01	0.186
500hPa风向	-0.13	2.435	0.15	2.818	0.17	3.204
400hPa风向	-0.17	3.204	0.13	2.435	0.10	1.867
700hPa风速	0	0	-0.04	0.744	-0.03	0.557
500hPa风速	-0.40	8.106	0.19	3.595	0.31	6.056
400hPa风速 ( $F_{400}$ )	-0.45	9.360	0.33	<b>6.493</b>	0.37	7.397

表2 不同大气污染物的预报气象因子

污染物浓度	预报气象因子
SO <sub>2</sub>	$\bar{T}$ 、 $T_{\text{max}}$ 、 $T_{\text{min}}$ 、 $T_{700}$ 、 $T_{500}$ 、 $T_{400}$ 、 $T_{d700}$ 、 $T_{d500}$
NO <sub>2</sub>	$\bar{T}$ 、 $T_{\text{min}}$ 、 $\bar{P}_{24}$ 、N、N <sub>1</sub> 、 $T_{d700}$ 、 $F_{400}$
PM <sub>10</sub>	$\bar{T}$ 、 $T_{\text{min}}$ 、U、N、 $T_{700}$ 、 $T_{500}$ 、 $T_{400}$ 、 $T_{d700}$

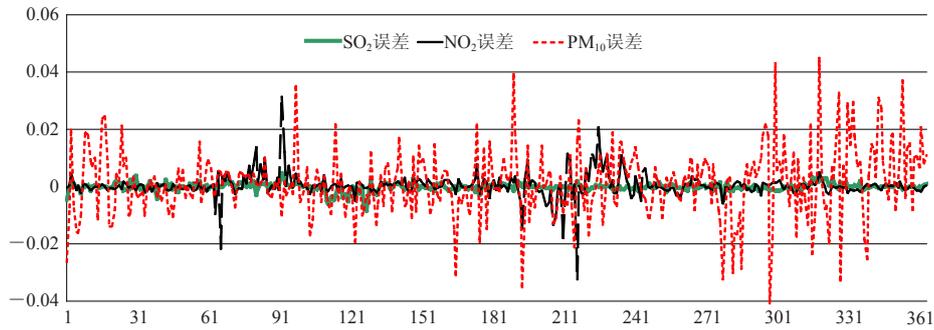


图2 林芝地区八一镇2011年5月1日—2012年4月30日动态空气质量预报的误差分析  
(横坐标代表该时段第1—第361天)

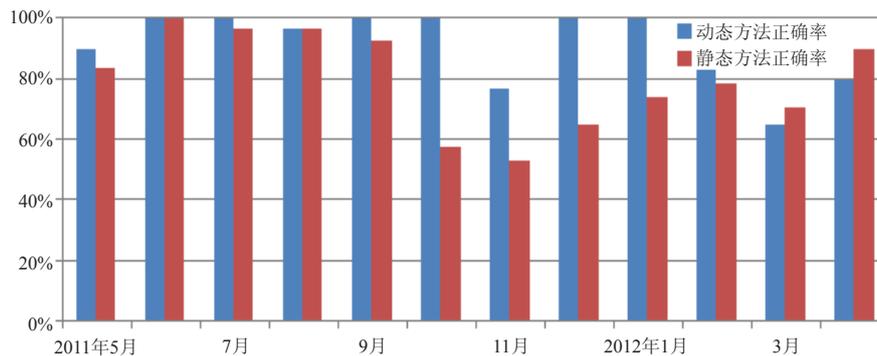


图3 林芝地区八一镇2011年5月1日—2012年4月30日动态空气质量预报的精度评价

$$Y_{31(\text{PM}_{10})} = -0.14055 + 0.011763\bar{T} - 0.00459T_{\min} + 0.001156U + 0.000244N - 0.00477T_{700} - 0.00085T_{500} - 0.00118T_{400} - 0.00245T_{d700} \quad (8)$$

将静态和动态计算方法得到的API值与实测API值进行对比,统计正确预报率,可以发现动态空气质量预报方法做出的平均正确率为91%,而静态空气质量预报方法做出的平均正确率为80%(图3)。

对动态空气质量预报方法和静态空气质量预报方法做置信度检验,发现动态空气质量预报方法预报出的API与实况API相关性达到0.567,通过0.01显著性水平的检验;而静态空气质量预报方法预报出相关性仅0.083,不能通过0.05显著性水平的检验。因此,动态空气质量预报方法由于充分考虑了气象因子与污染物的动态关系,相比静态方法具有更高的预报精度,具有一定的应用价值。

## 5 总结与讨论

通过对林芝地区环境空气质量预报方法的研究,发现动态建立回归方程并引入一些对影响污染物浓度变化有显著意义的气象因子,通过数理统计方法构造

数学模型是可行的。与静态方法对比,API的预报正确率相对较高,其原因在于污染物浓度与气象资料并非线性关系,长时期使用相对固定的预报方程导致误差不断累积。

由于污染物浓度与气象资料关系非常复杂,动态预报方法中时间窗大小的选择非常困难,本文初步地将20d作为统计时段,并未深入对时间窗大小的选择进行研究;同时,由于使用气象数据来模拟空气污染指数起步较晚,环境监测站点少,观测资料有限。因此,动态空气质量预报方法需要进一步研究和完善,才能形成一套完善可靠的动态预报方法。

### 参考文献

- [1] 朱玉强. 几种空气质量预报方法的预报效果对比分析. 气象, 2004, 30(10): 30-33.
- [2] 吴嘉荣. 用线性回归法建立城市环境空气质量预报模式. 引进与咨询, 2005, 12: 27-28.
- [3] 刘雨林. 西藏生态可持续发展研究. 黑龙江民族丛刊, 2007, 4: 60-68.
- [4] 旺杰, 周洪莉, 陈官燕, 等. 林芝气象与防灾减灾. 北京: 气象出版社, 2007.
- [5] 申建华. 浅析气象条件对大气污染时空分布的影响. 科技情报开发与经济, 2008, 18(25): 135-136.