

基于 Bayes 判别分析的大气污染潜势 定量预报*

周须文 高旭旭 于长文 韩世茹 许启慧

河北省气候中心, 河北省气象与生态环境重点实验室, 石家庄 050021

摘要: 本文依据大气污染物质量守恒方程, 构建了大气污染气象因子, 并以空气质量指数日增量为对象, 对大气污染潜势进行了量化分级划分。采用 Q 型聚类分析方法, 把秋冬季大气环流背景分为冷空气型、混合型、暖空气型三种环流类型, 并研究了区分三种大气环流类型的指标因子及其阈值。采用 Bayes 判别分析方法, 分别建立了冷空气型、混合型和暖空气型大气环流背景的污染潜势五级预测模型。对邢台 2017-2019 年秋冬季资料建立的预测模型, 判别正确率分别为 80%、71%、74.7%, 综合正确率 75.2%。采用 2015-2017 年秋冬季资料对该模型进行检验, 综合正确率为 63.6%。对 2019-2021 年秋冬季大气自净指数和模型判别结果与 AQI 日增量实况进行对比分析表明: 模型判别结果与 AQI 日增量的变化趋势有较高的一致性, 相关系数在 0.67 以上, 明显好于大气自净能力指数计算结果; 预测模型对极利于扩散等级和极不利于扩散等级判别正确次数明显高于大气自净能力指数。

关键词: 大气分类, 污染潜势, 预测模型, 数据挖掘技术

Prediction Method of Air Pollution Potential Based on Bayes Discriminant Analysis

ZHOU Xuwen GAO Xuxu YU Changwen HAN Shiru XU Qihui

Hebei Climate Center, Key Laboratory of Meteorology and Ecological Environment of Hebei Province, Shijiazhuang 050021, China

Abstract: Based on the equation of atmospheric pollutants mass conservation, the meteorological factors which affected on the air pollution were constructed and daily increment of air quality index (AQI) was as the object to quantitatively classify the air pollution potential. The atmospheric circulation background was divided into cold, mixed and warm air circulations in autumn and winter by the Q-type cluster analysis method. Then the meteorological factor and its thresholds for distinguishing three types of atmospheric circulation were studied. Used autumn-winter atmospheric data during 2017-2019, the five grads prediction models of pollution potential were established which the discriminant accuracy were 80%, 71% and 74.7% for cold, mixed and warm air circulations respectively, by the Bayes discriminant analysis method. The mean accuracy of three air types reached 75.2%, which can reach 63.6% when the five grads prediction models were tested by autumn-winter atmospheric data during 2015-2017. Through comparison of atmospheric self-purification index (ASI) and model discriminant results with daily increment of AQI in autumn and winter from 2019 to 2021, the results shows the model discriminant results was more consistent with the daily increment of AQI in trend than ASI, and the correlation coefficient was above 0.67. The predictably correct times of grad discrimination models in extremely favorable and extremely unfavorable pollutant diffusion were significantly more than by AQI.

Key word: atmospheric classification, air pollution potential, prediction model, data mining technology

引言

2013 年 1 月我国中东部地区集中出现雾-霾天气以来, 雾-霾天气一直是人们关注的焦点、研究的热点和政府治理大气污染的重点。众所周知, 治理大气污染的主要措施是减少向

*资助课题: 河北省科技计划项目 (17275303D) 和河北省气象局科研开发项目 (19kyd01) 共同资助

作者简介: 周须文, 主要从事短期气候预测。E-mail: hbszxwgood@163.com

大气中污染物的排放。这种措施必然会对经济发展产生一定影响，如何协调经济的持续发展
和大气环境保护是政府亟待解决的难题。研究表明，雾-霾天气形成的条件主要有两个：一
是人类活动向大气中排放的污染物，为霾的形成提供物质条件；二是静稳的气象条件。在污
染源基本稳定的前提下，空气质量的优劣主要由气象条件决定。开展大气污染潜势预报，为
防治大气污染提供科技支撑，做到精准治理雾-霾是化解这一难题的重要方法。

丁一汇和柳艳菊（2014）、江琪等（2017）、蔡子颖等（2017）、尚可等（2016）、于
文金等（2016）、梅梅等（2019）等专家学者对气象条件与空气污染关系的研究表明：空气
污染浓度与边界层内多个气象要素显著相关，引起空气污染浓度日均值波动因素的 70%是气
象条件变化。在环境气象服务中需要建立综合表达空气污染气象条件的量化指标，用于预测
和评估污染气象条件的长、中、短期变化。国外学者 Wang and Angell(1999)通过对日平均
海平面气压、地转风速、500 hPa 日平均水平风速、850 hPa 及以下逆温情况等设定判别条
件定义了静稳天气指数；Tai et al（2010）基于上述静稳天气的定义，统计分析了美国在
大气静稳日和非静稳日 $PM_{2.5}$ 浓度的差别。该静稳天气指数给出的是大气静稳的定性判断依
据，无法满足我国环境气象业务对定量描述大气静稳程度的需求。国内学者在对雾-霾天气
气象条件的定量描述上开展了许多指数的研发工作，朱蓉等（2001）基于箱模式理论建立了
空气污染潜势指数，并开发了城市大气污染数值预报系统（CAPPS）；杨元琴等（2009）选
取敏感气象要素：气温、气压、风、相对湿度、稳定度，并引入凝结函数开发了空气质量气
象参数（PLAM）指数；廖碧婷等（2012）利用 K 指数、沙氏指数和 L 指数构建了垂直交换系
数，对空气污染物的垂直输送能力进行了评估，并尝试对灰霾天气和无视程障碍天气进行预
报；徐大海等（2016）等提出了用 A 值法确定大气环境容量的方法；张恒德等（2017）根据预
报经验和历史气象要素统计，挑选与大气污染有关的气象要素及其阈值，通过权重求和初步
构建了静稳天气综合指数（SWI）；朱蓉等（2018）根据大气自身所具有的通风稀释和湿清
除能力，定义大气自净能力指数，给出基于常规气象观测的计算方法。上述指数从不同角度
体现了气象条件对空气污染的作用，但在环境气象业务应用中体现气象条件对大气污染的综合影响还有一定难度，本文结合气象业务部门现有数据环境，以大气污染物质量守恒方程为
理论基础，构建物理意义明确的气象因子，并选取与大气污染密切相关的气象要素，基于数
据挖掘技术建立大气污染潜势分级预测模型，实现大气污染潜势的客观化、定量化、精细化
预报。

1 资料与方法

1.1 资料

选用邢台 2015-2021 年秋冬季（10 月 1 日-翌年 3 月 31 日）地面、高空气象资料和空气
质量指数资料。地面气象资料包括逐日 4 个标准时次的资料（气温、气压、风速）及日资
料（平均气温、平均气压、相对湿度、平均风速和降水量）；高空气象资料为逐日 4 个标准
时次 1000 hPa、925 hPa、850 hPa、700 hPa、600 hPa 和 500 hPa 的气温、位势高度、风
速。其中地面气象资料和 08 时、20 时的高空气象资料由河北省气象信息中心提供；02 时
的高空资料由前一日 20 时和当日 08 时的相应资料计算平均所得，14 时的高空资料由当日 08
时和 20 时的相应资料计算平均所得。地面、1000 hPa、925 hPa 和 850 hPa 日平均垂直速
度，为距离邢台站最近格点（37.5° N，115.0° E）再分析资料，由 NCEP/NCAR 提供的。

（<https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html>）

邢台市 2015-2021 年秋冬季（10 月 1 日-翌年 3 月 31 日）逐日空气质量指数（AQI）资料来源于河北省生态环境监测中心空气质量实时发布平台。

1.2 方法

主要采用了统计分析方法中的聚类分析（施能，2009）和 Bayes 判别分析（Friedman et al, 1997）方法。

聚类分析是把一组数据按照相似性和差异性分为几个类别，使得属于同一类别的数据间的相似性尽可能大，不同类别数据间的相似性尽可能小。其中类的数量由系统按照某种性能指标自动生成。大气背景场的分类是对大气背景场个案的分类，分类数是未知的，因此采用聚类分析中的 Q 型聚类方法，以组间联接，夹角余弦法为相似性统计量，由系统按照组间平均距离自动生成多个数量的分类，根据不同类的特征来确定最佳分类数量。

判别分析是在分类确定的条件下，根据某一研究对象的各种特征值判别其类型归属问题的一种多变量统计分析方法。其基本原理是按照一定的判别准则，建立一个或多个判别函数，用研究对象的大量资料确定判别函数中的待定系数，并计算判别指标。据此可确定某一样本属于何类。大气污染潜势分五级进行预测，属于分类确定的多组别判别问题，因此选用最大优势是解决多组别判别问题的 Bayes 判别方法。

2 大气背景场的分类

杨旭等（2017）对京津冀地区冬半年空气污染天气进行了分型研究，表明不同天气类型下影响污染物扩散、输送、生成和转化的气象条件有所差异。周须文等（2020）对京津冀地区冬半年雾霾消散大气进行了分类，表明在不同类型大气下，对雾霾消散的影响因子及其贡献也不相同。因此有必要先对大气背景场进行分类研究。

2.1 大气背景场的聚类分析

对流层是大气圈最靠近地面的一层，受太阳辐射与大气环流的影响，对流层中下层的气流运动对天气现象的发生及大气污染物的扩散、输送和转化影响最大。对流层大气的气流运动状态可由各气压层气团热力性质变化来表征，因此选取对流层中下层标准气压层（500 hPa、600 hPa、700 hPa、850 hPa、925 hPa、1000 hPa）08 时的 24 小时变温和代表整个气柱冷暖气团变化的地面 24 小时变压作为研究大气背景场因子。

对邢台 2017-2019 年秋冬季逐日大气 362 个个案进行 Q 型聚类，分析自动生成的不同数量的分类个案特征，认为把大气背景场分三类的个案特征具有明确的天气学意义。图 1 是大气分三类中各类个案因子（24 小时变温）合成图，可以清楚看出三类大气特征具有明显差异：第 1 类对流层中下层（500 hPa 以下）均是明显的负变温；第 3 类均是正变温；第 2 类在边界层内（850 hPa 以下）为正变温，边界层以上（850 hPa~500 hPa）为负变温。表明第 1 类大气对流层中下层冷气团移入为主，称冷空气型；第 2 类大气在边界层内以正变温为主，边界层以上以负变温为主，对流层中下层呈上冷下暖的分布特征，称混合型大气；第 3 类大气对流层中下层均为正变温，说明多为暖气团移入，称暖空气型。显然该三类大气清楚地表明了边界层上部自由大气的运动状态。

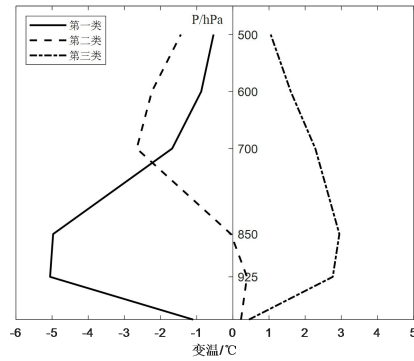


图 1 2017-2019 年秋、冬季邢台个例合成的三种大气环流类型

Fig1 Vertical distributions of meteorological factor for three atmospheric types in Xingtai during autumn and winter from 2017 to 2019

2.2 大气背景场分类的指标气象因子及阈值

在实际应用中如何识别和区分大气背景类型呢？还需要寻找能较好代表三种类型大气的指标气象因子及其阈值。

(1) 大气分类的指标气象因子分析

为了比较各气压层24小时变温因子在三类大气中的分布差异，对三类大气中各因子 ≤ 0 的个案数占该类总个案数的百分比进行了统计。根据三类大气型的天气学意义，冷空气型大气中，因子 ≤ 0 的个案数占比越大越好；暖空气型大气中，因子 ≤ 0 的个案数占比越小越好。由图2看到：冷空气型中850 hPa的24小时变温因子 ≤ 0 的个案数占比最大，为81.6%；暖空气型大气中850 hPa的24小时变温因子 ≤ 0 的个案数占比最少，只有2.5%；表明850 hPa的24小时变温因子在冷空气型和暖空气型大气中重叠个案最少。混合型大气中850 hPa的24小时变温因子 ≤ 0 的个案数占比为50%，说明该因子中位数为0，其值大多在0附近。从各类型大气因子平均值（图1）看到：在850 hPa层上，混合型大气因子平均值接近0，而冷空气型和暖空气型因子平均值与混合型大气因子平均值相差都较大，显然850 hPa的24小时变温因子较其他气压层的因子对三类大气更具代表性。因此采用850 hPa的24小时变温作为大气背景分类的指标气象因子。

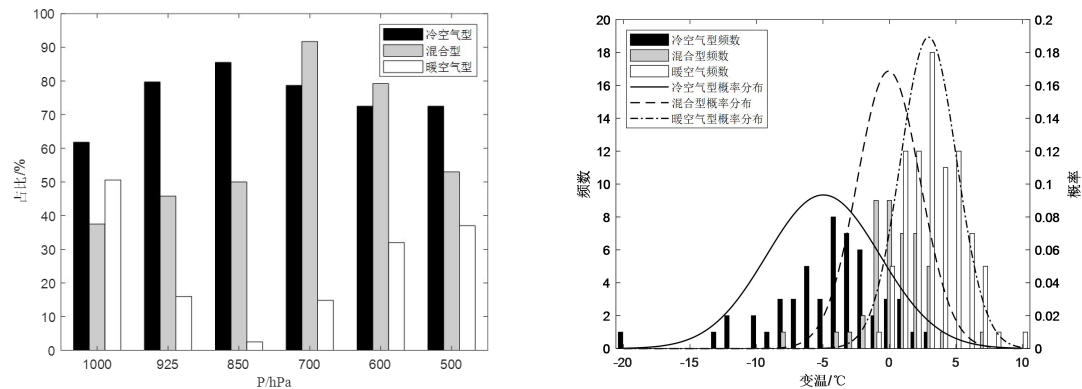


图2 2017-2019 年秋、冬季邢台三种大气环流类型因子 ≤ 0 的个例占比

Fig2 The case proportions of meteorological factor in different isopiestic surfaces for three atmospheric types in Xingtai during autumn and winter from 2017 to 2019.

图3 2017-2019年秋、冬季邢台指标因子在三种大气环流类型中的频数和概率分布

Fig3 Frequency and probability distribution of index factor for three atmospheric types in Xingtai during autumn and winter from 2017 to 2019.

(2) 指标因子阈值的确定

统计 850 hPa 的 24 小时变温因子在各类大气中的频数及概率分布（图 3），看到三类大气背景下的因子发生频次均呈现单峰形态，并且冷空气型、混合型和暖空气型因子自左至右呈错峰分布，峰值位置具有明显的差异。它们的概率分布都具有正态分布特征，相邻两类之间呈双峰结构，采用双峰法选择阈值，即在双峰之间的最低谷处就是两类分割的阈值所在，根据这一原理可以计算相近两类大气的分割阈值。计算结果：冷空气型与混合型大气分割阈值为 $-2.84\text{ }^{\circ}\text{C}$ ，混合型与暖空气型大气分割阈值 $1.36\text{ }^{\circ}\text{C}$ 。根据上述阈值对邢台 2017-2019 年秋冬季逐日大气进行区分，区分三类大气的正确百分率分别为：冷空气型 69%、混合型 61%、暖空气型 78%。显然选取的指标因子及其阈值对大气背景具有较好的分类效果。

3 大气污染潜势气象因子的选取

3.1 大气污染潜势气象因子构建的理论基础

根据大气污染物质量守恒方程（龚强等，1999）：

$$\frac{\partial C}{\partial t} = -V \cdot \nabla_2 C - \omega \frac{\partial C}{\partial p} + \nabla_3 (K \nabla_3 C) + Q + H \quad (1)$$

式中： C 为大气中的污染物浓度， V 为水平风速， ω 为垂直速度， P 为气压， K 为湍流扩散系数， Q 为污染物的排放量， H 为湿沉降量， ∇_2 代表水平二维向量算子， ∇_3 代表空间三维向量算子。

从方程（1）中看到，某地污染浓度变化取决于平流扩散、对流扩散、湍流扩散、污染物排放和湿沉降五大项。假设污染物排放量没有变化，污染浓度变化主要由其余四项决定。大气低层污染物的扩散、清除主要发生在混合层内，那么大气对污染物的平流扩散能力可用混合层的大气通风量表示：

$$V_E = \int_0^h u(z) dz \quad (2)$$

式中： V_E 为大气通风量， $u(z)$ 为混合层内 z 高度的风速， h 为混合层高度。

对流扩散是大气垂直运动的结果，因此对流扩散能力由大气边界层内的垂直速度 ω 来表示。

湍流扩散是大气中不规则、无组织的时空尺度跨度很大的气团运动，分为机械湍流和热力湍流。机械湍流激发的扩散能力可用风速的垂直切变率来表示：

$$V_R = \frac{U_2 - U_1}{Z_2 - Z_1} \quad (3)$$

式中： V_R 表示风速的垂直切变率， Z_2 、 Z_1 表示两个气压层的高度， U_2 、 U_1 表示两个气压层的风速。

热力湍流主要是在垂直方向产生温度梯度而造成的。国际原子能机构在 1980 年推出了

温度梯度法,即用两层大气间铅直方向的温度梯度来表示垂直方向上的湍流状态,判据可表示为:

$$T_R = \frac{\Delta T}{\Delta Z} \quad (4)$$

式中: T_R 表示铅直方向温度梯度, ΔT 表示两个气压层的温度差, ΔZ 表示两个气压层的厚度。

湿沉降是指降水对大气中污染物冲刷使得大气中污染物浓度降低。湿沉降能力可表示为:

$$V_w = W_r \cdot R \quad (5)$$

式中: W_r 表示雨洗常数,取 6×10^5 , R 为降水率。

3.2 气象因子的计算方法

(1) 混合层高度

大气边界层的上限高度即为混合层高度,因天气条件而异,白天与夜晚不同,城乡也有明显差别。在中纬度城市,通常晴天的白天可达到 1000~1500 m (850 hPa), 夜晚只有 200 m 左右。由于受观测资料所限,把 850 hPa 以下的大气认为是大气边界层,由于大气边界层内不同高度的气象因子对污染物迁移扩散作用不同,根据气象部门现有的气象资料,把混合层高度按标准气压层 (850 hPa、925 hPa) 分为两段 (地面~925 hPa、925 hPa~850 hPa) 分别计算。

根据天气学原理,两等压面之间的厚度计算公式为:

$$h = R \cdot T_m \cdot \ln \left(\frac{P_1}{P_0} \right) / 9.8$$

其中: 比气体常数 $R=2.87$, $T_m = 0.5(T_1 + T_0)$ (等压面 P_1 和 P_0 气温平均值)。

地面~925 hPa 的厚度记为 h_1 , 925 hPa~850 hPa 的厚度记为 h_2 。

混合层高度: $h = h_1 + h_2$

(2) 平流扩散因子—大气通风量:

地面~925 hPa 的通风量: $V_{E1} = 0.5(u_{地面} + u_{925}) \cdot h_1$

925 hPa~850 hPa 的通风量: $V_{E2} = 0.5(u_{925} + u_{850}) \cdot h_2$

(3) 对流扩散因子—垂直速度:

选取了地面 ($\omega_{地}$) 和边界层内标准气压层的垂直速度 (ω_{1000} 、 ω_{925} 、 ω_{850})。

(4) 机械湍流因子—风速垂直切变:

地面~925 hPa 的风切变: $V_{R1} = (u_{925} - u_{地面})/h_1$

925 hPa~850 hPa 的风切变: $V_{R2} = (u_{850} - u_{925})/h_2$

(5) 热力湍流因子—垂直温度梯度:

地面~925 hPa 的温度梯度: $T_{R1} = (T_{925} - T_{地面})/h_1$

925 hPa~850 hPa 的温度梯度: $T_{R2} = (T_{850} - T_{925})/h_2$

(6) 湿沉降因子代表的是降水对大气中污染物的冲刷能力,因此选取 24 小时降水量作为湿沉降因子。

3.3 大气污染潜势气象因子的选取

根据实际气象资料情况,选取地面和边界层内标准气压层的日平均垂直速度作为对流扩

散因子；选取 24 小时降水量作为湿沉降因子，因子对应符号见表 1a。平流扩散因子选取大气边界层内两个高度层 4 个时次的大气通风量，共 8 个因子；湍流扩散因子选取大气边界层内两个高度层 4 个时次的水平风速切变、垂直温度梯度，共 16 个因子。因子对应符号见表 1b。另外，根据杨元琴等（2009）、张恒德等（2017）的研究成果，选取了地面敏感气象要素：日平均气温、日平均气压、日平均相对湿度、日平均风速、24 小时变压、24 小时变温，因子对应符号见表 1c。

表 1a 对流扩散、湿沉降因子名称、符号对照表

Tab 1a. Reference table of atmospheric factor symbols for convection diffusion and wet depositions

名称	地面垂直速度	1000 hPa 垂直速度	925 hPa 垂直速度	850 hPa 垂直速度	24 小时降水量
符号	$\omega_{地}$	ω_{1000}	ω_{925}	ω_{850}	R_{24}

表 1b 平流扩散、湍流扩散因子名称、符号对照表

Tab 1b Reference table of atmospheric factor symbols for advection and turbulent diffusion

因子名称	高度层	时次			
		02 时	08 时	14 时	20 时
大气通风量	地面~925 hPa	V_{E102}	V_{E108}	V_{E114}	V_{E120}
	925hPa~850 hPa	V_{E202}	V_{E208}	V_{E214}	V_{E220}
垂直温度梯度	地面~925h Pa	T_{R102}	T_{R108}	T_{R114}	T_{R120}
	925 hPa~850 hPa	T_{R202}	T_{R208}	T_{R214}	T_{R220}
风速垂直切变率	地面~925 hPa	V_{R102}	V_{R108}	V_{R114}	V_{R120}
	925 hPa~850 hPa	V_{R202}	V_{R208}	V_{R214}	V_{R220}

表 1c 地面气象因子、符号对照表

Tab 1c. Reference table of meteorological factor symbols in surface

名称	24 小时变压	24 小时变温	日平均气温	日平均气压	日平均相对湿度	日平均风速
符号	ΔP_{24}	ΔT_{24}	T	P	rh	fs

4 大气污染潜势分级预测模型建立

4.1 大气污染潜势分级划分

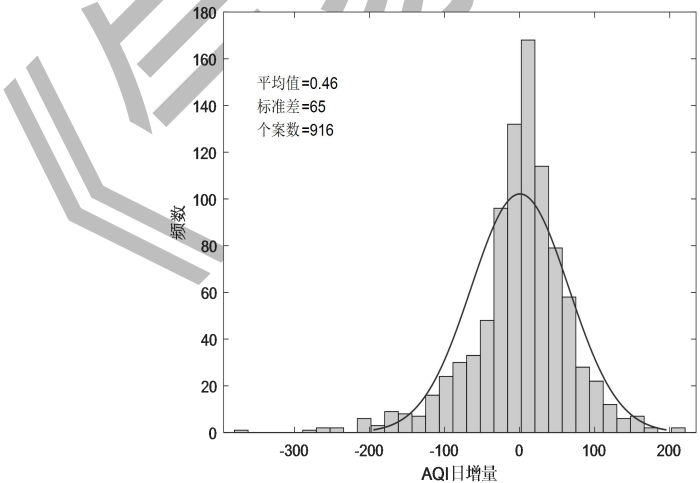


图 4 2015-2019 年秋、冬季邢台 AQI 日增量分布

Fig4 Frequency distribution of AQI daily increment in Xingtai during autumn and winter from 2015 to 2019.

大气污染潜势直接影响空气质量的变化，因此大气污染潜势的分级可依据空气质量指数

(AQI) 日增量来划分。对邢台 2015–2019 年秋冬季 AQI 日增量统计分析 (图 4): AQI 日增量出现频率呈正态分布, AQI 日增量平均值为 0.46, 标准差为 65, 认为大气污染潜势五级划分比较合理; 参考全国气象防灾减灾标准化技术委员会 (2018), 把 AQI 日增量-80、-30、30、80 作为分界值 (表 2)。划分的 AQI 日增量五分级与 AQI 日增量的相关性为 0.95, 说明 AQI 日增量五分级能较好代表 AQI 日变化状况。

表 2 污染潜势五级等级标准和描述
Tab 2. Five grades of air pollution potential

等级	AQI 日增量	描述
1	≤ -80	气象条件非常有利于空气污染物稀释、扩散和清除
2	$-80 \sim -30$	气象条件有利于空气污染物稀释、扩散和清除
3	$-30 \sim 30$	气象条件对空气污染物稀释、扩散和清除无明显影响
4	$30 \sim 80$	气象条件不利于空气污染物稀释、扩散和清除
5	$80 \geq$	气象条件非常不利于空气污染物稀释、扩散和清除

4.2 污染潜势五分级预测模型建立

本文创建预测模型采用的是 2017–2019 年秋冬季邢台的资料。首先对建模资料进行处理: (1) 根据大气背景分类指标因子及其阈值对研究的个案进行识别分类, 建立冷空气型、混合型和暖空气型三类大气影响因子数据库。(2) 对各类型大气个案的 AQI 日增量按表 2 的污染潜势五分级标准进行划分, 构建带有标号的模型创建训练数据集。

由于选取的变量因子较多, 在应用 Bayes 判别方法建立判别函数过程中, 通过逐步选择法对自变量因子进行筛选, 剔出了最低容差水平在 0.001 以下的变量, 最后保留了对模型判别贡献较大的变量因子。通过对数据集训练确定了五分级判别函数的变量因子系数 (见表 3、表 4、表 5)。对比表 3、表 4、表 5 发现, 不同大气背景的五分级判别函数, 参与的变量因子及其权重都有所不同, 说明大气中污染物的聚集和消散是复杂多变的, 不同大气背景下它的变化机制存在较大差异, 分类建立预测模型是必要的。

表 3 冷空气型大气污染潜势五分级判别函数系数
Tab 3 Five grades discriminant function of pollution potential for cold air type

变量因子符号	变量编号	一级	二级	三级	四级	五级
ΔP_{24}	X_1	1.57	1.438	1.232	1.587	2.219
ΔT_{24}	X_2	-0.611	-0.75	-0.475	-0.225	1.136
rh	X_3	0.547	0.558	0.533	0.58	0.391
fs	X_4	5.886	6.435	5.618	5.475	3.353
V_{E102}	X_5	0.064	0.057	0.045	0.064	0.078
V_{E108}	X_6	0.027	0.019	0.02	0.026	0
V_{E141}	X_7	-0.085	-0.076	-0.064	-0.09	-0.099
V_{E202}	X_9	-0.055	-0.05	-0.039	-0.055	-0.067
V_{E208}	X_{10}	-0.031	-0.023	-0.023	-0.03	-0.007
V_{E214}	X_{11}	0.09	0.08	0.068	0.095	0.101
T_{R102}	X_{12}	11.545	11.322	4.901	8.999	39.136
T_{R108}	X_{13}	24.352	26.816	29.293	31.485	10.229
T_{R114}	X_{14}	-48.602	-50.867	-44.598	-51.804	-63.095
T_{R202}	X_{15}	-9.236	-10.223	-12.488	-10.711	-19.838
T_{R208}	X_{16}	-11.778	-8.168	-5.91	-4.901	-24.892

T_{R214}	X_{17}	-18.647	-20.104	-15.585	-18.594	3.473
V_{R102}	X_{18}	113.96	102.322	80.267	106.541	122.395
V_{R108}	X_{19}	74.001	57.353	53.149	76.409	28.08
V_{R114}	X_{20}	-211.39	-190.612	-157.361	-224.287	-239.453
V_{R202}	X_{22}	169.279	154.69	125.881	176.865	188.553
V_{R208}	X_{23}	47.454	32.032	33.865	40.092	-10.007
V_{R214}	X_{24}	-184.127	-163.631	-137.354	-195.532	-210.011
$\omega_{\text{地}}$	X_{25}	-64.508	-52.518	-57.87	-72.354	7.676
ω_{1000}	X_{26}	292.146	260.469	230.189	249.041	246.443
ω_{925}	X_{27}	-276.66	-231.071	-164.605	-170.643	-277.51
ω_{850}	X_{28}	79.364	58.711	27.415	36.046	85.526
常量		-64.292	-59.405	-51.758	-64.256	-65.768

表 4 混合型大气污染潜势五分级判别函数系数

Tab 4 Five grades discriminant function of pollution potential for mixed air type

变量因子符号	变量编号	一级	二级	三级	四级	五级
ΔT_{24}	X_2	0.43	0.437	0.597	0.741	1.031
rh	X_3	0.524	0.553	0.541	0.541	0.56
fs	X_4	6.59	6.34	6.208	6.898	5.058
V_{E108}	X_6	0.009	0.008	0.009	0.009	0.009
V_{E120}	X_8	0.011	0.01	0.011	0.011	0.01
V_{E202}	X_9	0.007	0.008	0.008	0.007	0.006
V_{E214}	X_{11}	-0.018	-0.015	-0.018	-0.017	-0.016
T_{R102}	X_{12}	-21.745	-23.46	-25.537	-26.613	-34.488
T_{R108}	X_{13}	14.641	16.468	19.199	21.001	29.377
T_{R202}	X_{15}	-21.097	-21.214	-22.573	-20.78	-18.405
T_{R208}	X_{16}	2.16	4.081	4.182	4.517	-0.342
T_{R214}	X_{17}	6.112	1.528	3.217	4.492	7.484
V_{R102}	X_{18}	-12.879	-18.565	-19.349	-19.921	-17.612
V_{R108}	X_{19}	-11.53	-12.555	-11.737	-11.113	-12.573
V_{R114}	X_{20}	49.365	42.562	51.237	51.201	49.045
V_{R208}	X_{23}	-9.54	-11.03	-9	-8.223	-5.428
V_{R214}	X_{24}	57.61	51.459	56.604	54.727	50.149
$\omega_{\text{地}}$	X_{25}	-39.443	-52.56	-44.059	-52.265	-40.142
ω_{1000}	X_{26}	169.159	156.023	115.224	141.775	75.431
ω_{925}	X_{27}	-100.697	-89.713	-59.584	-78.901	-12.873
常量		-45.264	-49.302	-44.399	-45.588	-46.956

表 5 暖空气型大气污染潜势五分级判别函数系数

Tab 5 Five grades discriminant function of pollution potential for warm air type

变量因子符号	变量编号	一级	二级	三级	四级	五级
ΔP_{24}	X_1	0.327	0.195	-0.085	0.017	-0.143
ΔT_{24}	X_2	1.137	1.042	0.782	0.95	0.858

<i>rh</i>	X ₃	0.523	0.455	0.417	0.446	0.464
<i>fs</i>	X ₄	8.29	5.067	7.353	6.943	5.637
<i>V_{E102}</i>	X ₅	0.02	0.019	0.019	0.02	0.019
<i>V_{E108}</i>	X ₆	-4.58E-05	0	0.001	7.60E-05	0.001
<i>V_{E141}</i>	X ₇	-0.001	0.002	0.001	0.001	0.001
<i>V_{E202}</i>	X ₉	-0.012	-0.012	-0.015	-0.015	-0.013
<i>T_{R102}</i>	X ₁₂	-22.681	-36.564	-37.418	-36.836	-39.574
<i>T_{R108}</i>	X ₁₃	16.088	22.316	27.232	27.51	27.107
<i>T_{R114}</i>	X ₁₄	5.775	8.62	-0.573	0.381	4.061
<i>T_{R202}</i>	X ₁₅	-16.996	-8.058	-23.245	-18.938	-12.29
<i>T_{R208}</i>	X ₁₆	21.002	3.993	15.15	12.843	8.783
<i>V_{R102}</i>	X ₁₈	32.977	25.345	29.852	30.648	27.804
<i>V_{R141}</i>	X ₂₀	-17.03	-3.309	-5.12	-7.058	-9.156
<i>V_{R120}</i>	X ₂₁	10.537	3.989	5.475	6.079	8.2
<i>V_{R202}</i>	X ₂₂	49.65	44.713	41.259	45.031	40.797
<i>V_{R208}</i>	X ₂₃	-11.828	-5.754	-6.013	-7.824	-7.019
<i>V_{R214}</i>	X ₂₄	-0.54	2.078	7.375	6.913	7.505
<i>ω_地</i>	X ₂₅	6.326	-32.471	-6.804	-16.775	-11.999
<i>ω₁₀₀₀</i>	X ₂₆	41.698	105.327	123.208	104.036	140.788
<i>ω₉₂₅</i>	X ₂₇	38.826	19.556	-181.341	-110.13	-157.027
<i>ω₈₅₀</i>	X ₂₈	-33.859	-56.697	114.497	66.265	78.047
常量		-52.297	-44.061	-41.398	-39.853	-43.337

把大气污染潜势1~5级判别函数的预测概率依次记为：P1、P2、P3、P4、P5，以冷空气型为例，给出了大气污染潜势五分级预测模型：

$$P_1 = -64.292 + 1.57X_1 - 0.611X_2 + 0.547X_3 + 5.886X_4 + 0.064X_5 + 0.027X_6 - 0.085X_7 - 0.055X_9 - 0.031X_{10} + 0.09X_{11} + 11.545X_{12} + 24.352X_{13} - 48.602X_{14} - 9.236X_{15} - 11.778X_{16} - 18.647X_{17} + 113.96X_{18} + 74.001X_{19} - 211.39X_{20} + 169.279X_{22} + 47.454X_{23} - 184.127X_{24} - 64.508X_{25} + 292.146X_{26} - 276.66X_{27} + 79.364X_{28}$$

$$P_2 = -59.405 + 1.44X_1 - 0.75X_2 + 0.56X_3 + 6.435X_4 + 0.057X_5 + 0.019X_6 - 0.076X_7 - 0.05X_9 - 0.02X_{10} + 0.08X_{11} + 11.322X_{12} + 26.816X_{13} - 50.867X_{14} - 10.223X_{15} - 8.168X_{16} - 20.104X_{17} + 102.322X_{18} + 57.35X_{19} - 190.612X_{20} + 154.69X_{22} + 32.032X_{23} - 163.631X_{24} - 52.518X_{25} + 260.469X_{26} - 231.071X_{27} + 58.711X_{28}$$

$$P_3 = -51.758 + 1.232X_1 - 0.475X_2 + 0.533X_3 + 5.618X_4 + 0.045X_5 + 0.02X_6 - 0.064X_7 - 0.039X_9 - 0.023X_{10} + 0.068X_{11} + 4.901X_{12} + 29.293X_{13} - 44.598X_{14} - 12.488X_{15} - 5.91X_{16} - 15.585X_{17} + 80.267X_{18} + 53.149X_{19} - 157.361X_{20} + 125.881X_{22} + 33.865X_{23} - 137.354X_{24} - 57.87X_{25} + 230.189X_{26} - 164.605X_{27} + 27.415X_{28}$$

$$P_4 = -64.256 + 1.587X_1 - 0.225X_2 + 0.58X_3 + 5.475X_4 + 0.064X_5 + 0.03X_6 - 0.09X_7 - 0.055X_9 - 0.03X_{10} + 0.095X_{11} + 8.999X_{12} + 31.485X_{13} - 51.804X_{14} - 10.711X_{15} - 4.901X_{16} - 18.594X_{17} + 106.541X_{18} + 76.41X_{19} - 24.287X_{20} + 176.865X_{22} + 40.092X_{23} - 195.532X_{24} - 72.354X_{25} + 249.041X_{26} - 170.643X_{27} + 36.046X_{28}$$

$$P_5 = -65.768 + 2.219X_1 + 1.136X_2 + 0.391X_3 + 3.353X_4 + 0.078X_5 - 0.099X_7 - 0.067X_9 - 0.007X_{10} + 0.101X_{11} + 39.136X_{12} + 10.229X_{13} - 63.095X_{14} - 19.838X_{15} - 24.892X_{16} + 3.473X_{17} + 122.395X_{18} + 28.08X_{19} - 239.453X_{20} + 188.553X_{22} - 10.007X_{23} - 210.011X_{24} + 7.676X_{25} + 246.443X_{26} - 277.51X_{27} + 85.526X_{28}$$

上述判别函数通过自身验证得到的大气污染潜势五级判别正确率为 75.2%，其中冷空气

型预测模型判别正确率为 80%。混合型预测模型判别正确率为 71%，暖空气型预测模型判别正确率为 74.7%。

4 大气污染潜势五分级预测模型检验

4.1 模型独立检验

模型独立检验法是对判别分析得到的判别函数效果评判最客观有效的方法。本文把 2015-2016 年秋冬季邢台逐日气象资料代入预测模型中，比较各级判别函数值（ P_1 、 P_2 、 P_3 、 P_4 、 P_5 ）的大小，根据 Bayes 判别准则，函数值最大代表所对应级别出现概率最大，即是预测级别。把污染潜势预测级别与实际 AQI 日增量级别进行对比，预测级别与 AQI 日增量级别相一致的记为正确，预测正确日数占实际出现日数的百分比即为预测正确率。表 6 为 2015-2016 年秋冬季邢台污染潜势五分级预测正确率，看到三类大气的预测模型虽然对各等级的污染潜势预测正确率存在差别，但对各等级正确率的分布形态基本一致。都对大气处于正常状态的三级预测最好，正确率在 80%以上；其次是大气非常利于扩散的一级和大气非常不利于扩散的五级，它们的综合正确率分别为 64.8%和 59.1%。该预测模型对 2015-2016 年秋冬季污染潜势五分级预测正确率 63.6%，能较好预测大气对污染物的聚集、扩散能力。

表 6 2015-2016 年秋冬季污染潜势五分级预测正确率

Tab 6 Five grades prediction accuracy of air pollution potential in autumn and winter from 2015 to 2016

	一级	二级	三级	四级	五级	综合正确率
冷空气型	69.6	58.8	81.6	33.3	100	70.6
混合型	54.5	33.3	83.9	29.4	75	56.1
暖空气型	66.7	50	80.5	68.5	52.9	64.5
各级正确率	64.8	45.2	82.2	57.1	59.1	63.6

4.2 与大气自净能力指数预测对比分析

4.2.1 大气自净能力指数（ASI）及其等级划分

大气自净能力指数是对雾-霾天气气象条件定量描述指数，它反映的是大气运动对污染物的通风稀释作用和降水对污染物的湿清除作用。根据朱蓉等（2018）提出的大气自净能力指数（ASI）计算方法，计算邢台 2019-2021 年秋冬季的大气自净能力指数。

为了便于与预测模型判别结果对比，把 ASI 也划分为五级进行评价。ASI 指数五级划分的依据：（1）将大气自净能力指数按 $2.5 \text{ t}/(\text{d} \cdot \text{km}^2)$ 和 $4.1 \text{ t}/(\text{d} \cdot \text{km}^2)$ 两个限值，分为了优、良和一般三个评价等级（全国气候与气候变化标准化技术委员会,2020）。（2）朱蓉等(2018)对大气自净能力指数的研究成果，当 ASI 低于 $1.4 \text{ t}/(\text{d} \cdot \text{km}^2)$ 时，京津冀地区容易出现 AQI 达到 200 的空气质量重污染等级；（3）结合 ASI 与 AQI 日增量的统计分析，当 ASI 低于 $0.4 \text{ t}/(\text{d} \cdot \text{km}^2)$ 时，AQI 日增量在 80 以上，ASI 的分级阈值和等级描述见表 7。对 ASI 划分的 5 等级与 ASI 相关系数为-0.915，通过了 $\alpha = 0.01$ 级别的相关性显著，表明该五分级能较好地代表大气的自净能力指数。

表 7 ASI 评价等级

Tab 7 Evaluation grads of ASI

ASI / $\text{t}/(\text{d} \cdot \text{km}^2)$	等级	描述
$\text{ASI} > 4.1$	1	极利于扩散
$2.5 < \text{ASI} \leq 4.1$	2	利于扩散

$1.4 < ASI \leq 2.5$	3	不明显
$0.4 < ASI \leq 1.4$	4	不利于扩散
$ASI \leq 0.4$	5	极不利于扩散

4.2.2 大气自净能力指数（ASI）和污染潜势五分级预测对比

对2019-2021年秋、冬季邢台逐日AQI日增量、ASI和污染潜势五分级预测进行了双变量相关分析（表8），可以看到：AQI日增量五分级与AQI日增量相关系数为0.947，表明AQI日增量五分级能较好地代表大气对物的稀释扩散能力。ASI五分级与AQI日增量、AQI日增量五分级的相关性均高于ASI，说明ASI五分级在对大气扩散能力描述上优于ASI；污染潜势五分级预测与AQI日增量、AQI日增量五分级的相关性分别为0.676和0.679，明显高于ASI和ASI五分级。显然预测模型的污染潜势五分级预测优于ASI。

表8 2019-2021年秋、冬季邢台AQI日增量、ASI和污染潜势五分级预测相关分析

Tab 8 The correlation coefficients between AQI daily increment, ASI and model prediction grads in Xingtai during autumn and winter from 2019 to 2021.

	AQI日增量	AQI日增量五分级	ASI	ASI五分级
AQI日增量五分级	0.947			
ASI	-0.436	-0.383		
ASI五分级	0.469	0.422	-0.915	
污染潜势五分级预测	0.676	0.679	-0.396	0.398

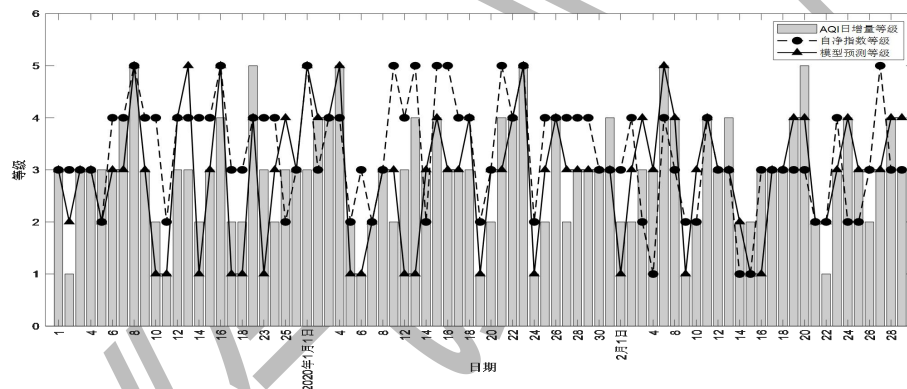


图5 2019年12月1日—2020年2月29日邢台AQI日增量、ASI、污染潜势预测等级时间序列

Fig 5. Time series of AQI daily increment, ASI and model prediction grads in Xingtai from 1 December 2019 to 29 February 2020.

图5为2019年12月1日—2020年2月29日邢台逐日AQI日增量等级实况和ASI等级、污染潜势预测等级，无论是正常生产时段还是疫情封闭时段（2020年1月24日-2月29日），AQI日增量等级变化没有明显差异，说明AQI日增量主要受大气条件的影响，污染物的排放量对其影响不大；模型判别的污染潜势等级与AQI日增量等级的变化趋势具有较高的一致性，明显好于ASI等级；模型对极利于扩散的一级和极不利于扩散的五级判别正确次数明显高于ASI；模型判别的污染潜势等级与AQI日增量等级实况基本上是相同或相差一级，而ASI等级与AQI日增量等级相差两级以上的较多。显然，模型对大气污染潜势的五级判别能力好于ASI。

5 总结与讨论

本文介绍了应用数据挖掘技术研究大气污染潜势分级预报方法,对邢台建立的大气污染潜势五分级预测模型进行了检验分析,得出结论如下:

(1) 通过对大气环流背景的分类研究,发现大气背景场可分为具有明确物理意义的三种大气类型(冷空气型、混合型、暖空气型),850 hPa的24小时变温是区分三类大气的最佳指标因子;采用双峰法选择的指标因子阈值能较好的区分三类大气,对冷空气型、混合型、暖空气型大气区分正确率分别为:69%、61%和78%。

(2) 应用数据挖掘技术客观选取变量因子及其权重,结果表明:大气中污染物的聚集和消散都是复杂的过程,有众多因子参与,不同大气背景各因子的贡献有较大差异。

(3) 建立的大气污染潜势五分级预测模型能较好地对AQI日增量进行五级判别,综合判别正确率为63.6%,其中冷空气型预测模型综合判别正确率最高,为70.6%。

(4) 大气污染潜势五分级预测模型的判别结果与实际大气AQI日增量等级的变化趋势具有较高的一致性,相关系数在0.67以上;明显好于大气自净指数对污染潜势的预测结果。

数据挖掘大多数是大而全,多而精,数据越多模型越可能精确,变量越多,数据之间的关系越明确。由于受现有气象资料限制,对气象因子的选用和计算都较粗,随着观测技术的进步,还需要在选取因子及其计算方面进行不断改进,使建立的预测模型准确率不断提高。

参考文献

- 蔡子颖,姚青,韩素芹,等,2017. 21世纪以来天津细颗粒物气象扩散能力趋势分析[J]. 中国环境科学, 37(6): 2040-2046.
- Cai Z Y, Yao Q, Han S Q, et al, 2017. The trends of fine particulate meteorological diffusivity in Tianjin from 21th[J]. China Environ Sci, 37(6): 2040-2046 (in Chinese).
- 丁一汇,柳艳菊,2014. 近50年我国雾和霾的长期变化特征及其与大气湿度的关系[J]. 中国科学:地球科学, 44(1): 37-48.
- Ding Y H, Liu Y J, 2014. Analysis of long-term variations of fog and haze in China in recent 50 years and their relations with atmospheric humidity[J]. Sci China Earth Sci, 57(1): 36-46.
- 龚强,白乐生,杨宏斌,1999. 空气污染潜势指数预报方法[J]. 气象与环境学报, (3): 46-47. Gong Q, Bai L S, Yang H B, 1999. Prediction method of air pollution potential index[J]. J Meteor Environ, (3): 46-47 (in Chinese).
- 江琪,王飞,张恒德,等,2017. 北京市PM_{2.5}和反应性气体浓度的变化特征及其与气象条件的关系[J]. 中国环境科学, 37(3): 829-837.
- Jiang Q, Wang F, Zhang H D, et al, 2017. Analysis of temporal variation characteristics and meteorological conditions of reactive gas and PM_{2.5} in Beijing[J]. China Environ Sci, 37(3): 829-837 (in Chinese).
- 廖碧婷,吴兑,陈静,等,2012. 灰霾天气变化特征及垂直交换系数的预报应用[J]. 热带气象学报, 28(3): 417-424. Liao B T, Wu D, Chen J, et al, 2012. A study of the variation trends of haze and application of the vertical exchange coefficient[J]. J Trop Meteor, 28(3): 417-424 (in Chinese).
- 梅梅,朱蓉,孙朝阳,2019. 京津冀及周边“2+26”城市秋冬季大气重污染气象条件及其气候特征研究[J]. 气候变化研究进展, 15(3): 270-281. Mei M, Zhu R, Sun C Y, 2019. Study on meteorological conditions for heavy air pollution and its climatic characteristics in “2+26” cities around Beijing-Tianjin-Hebei region in autumn and winter[J]. Climate Change Res, 15(3): 270-281 (in Chinese).
- 全国气候与气候变化标准化技术委员会,2020.气候资源评价 气候宜居城镇:QX/T 570-2020[S]. 北京:气象出版社,2020. Committee 540 on Climate and Climate Change of Standardization Administration of China, 2020. Climate resource assessment-Climate livable cities:QX/T 570-2020[S]. Beijing: China Meteorological Press, 2020.
- 全国气象防灾减灾标准化技术委员会,2018.空气污染扩散气象条件等级:QX/T 413-2018[S]. 北京:气象出版社,2018. Committee 345 on meteorological disaster prevention and mitigation of Standardization Administration of China, 2018. Grades of

- air pollution diffusion meteorological conditions:QX/T 413-2018[S]. Beijing: China Meteorological Press, 2018.
- 尚可, 杨晓亮, 张叶, 等, 2016. 河北省边界层气象要素与 $PM_{2.5}$ 关系的统计特征[J]. 环境科学研究, 29(3): 323-333. Shang K, Yang X L, Zhang Y, et al, 2016. Statistical analysis of the relationship between meteorological factors and $PM_{2.5}$ in the boundary layer over Hebei province[J]. Res Environ Sci, 29(3): 323-333 (in Chinese).
- 施能, 2009. 气象统计预报[M]. 北京: 气象出版社: 206-219. Shi N, 2009. Meteorological Statistical Forecast[M]. Beijing: Meteorological Press: 206-219 (in Chinese).
- 徐大海, 王郁, 朱蓉, 2016. 大气环境容量系数 A 值频率曲线拟合及其应用[J]. 中国环境科学, 36(10): 2913-2922. Xu D H, Wang Y, Zhu R, 2016. The atmospheric environmental capacity coefficient cumulative frequency curve fitting and its application[J]. China Environ Sci, 36(10): 2913-2922 (in Chinese).
- 杨旭, 张小玲, 康延臻, 等, 2017. 京津冀地区冬半年空气污染天气分型研究[J]. 中国环境科学, 37(9): 3201-3209. Yang X, Zhang X L, Kang Y Z, et al, 2017. Circulation weather type classification for air pollution over the Beijing-Tianjin-Hebei region during winter[J]. China Environ Sci, 37(9): 3201-3209 (in Chinese).
- 杨元琴, 王继志, 侯青, 等, 2009. 北京夏季空气质量的气象指数预报[J]. 应用气象学报, 20(6): 649-655. Yang Y Q, Wang J Z, Hou Q, et al, 2009. A PLAM index forecast method for air quality of Beijing in Summer[J]. J Appl Meteor Sci, 20(6): 649-655 (in Chinese).
- 于文金, 吴雁, 黄亦露, 等, 2016. 河北省雾霾波动变化特征及成因研究[J]. 大气科学学报, 39(4): 554-561. Yu W J, Wu Y, Huang Y L, et al, 2016. The variation and genesis of haze in Hebei Province[J]. Trans Atmos Sci, 39(4): 554-561 (in Chinese).
- 张恒德, 张碧辉, 吕梦瑶, 等, 2017. 北京地区静稳天气综合指数的初步构建及其在环境气象中的应用[J]. 气象, 43(8): 998-1004. Zhang H D, Zhang B H, Lü M Y, et al, 2017. Development and application of stable weather index of Beijing in environmental meteorology[J]. Meteor Mon, 43(8): 998-1004 (in Chinese).
- 周须文, 韩世茹, 井元元, 等, 2020. 京津冀雾霾消散因子及其阈值研究[J]. 气象学报, 78(1): 154-162. Zhou X W, Han S R, Jing Y Y, et al, 2020. A study on haze dissipation factors and their thresholds over Beijing-Tianjin-Hebei area[J]. Acta Meteor Sin, 78(1): 154-162 (in Chinese).
- 朱蓉, 徐大海, 孟燕君, 等, 2001. 城市空气污染数值预报系统 CAPPS 及其应用[J]. 应用气象学报, 12(3): 267-278. Zhu R, Xu D H, Meng Y J, et al, 2001. City air pollution numerical prediction system and its application[J]. Quart J Appl Meteor, 12(3): 267-278 (in Chinese).
- 朱蓉, 张存杰, 梅梅, 2018. 大气自净能力指数的气候特征与应用研究[J]. 中国环境科学, 38(10): 3601-3610. Zhu R, Zhang C J, Mei M, 2018. The climate characteristics of atmospheric self-cleaning ability index and its application in China[J]. China Environ Sci, 38(10): 3601-3610 (in Chinese).
- Friedman N, Geiger D, Goldszmidt M, 1997. Bayesian network classifier[J]. Mach Learn, 29(1): 131-163.
- Tai A P K, Mickley L J, Jacob D J, 2010. Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: implications for the sensitivity of $PM_{2.5}$ to climate change[J]. Atmos Environ, 44(32): 3976-3984.
- Wang J X L, Angell J K, 1999. Air stagnation climatology for the United States (1948-1998)[R]. Silver Spring, MD: National Oceanic and Atmospheric Administration.