



文章编号:1001-9081(2017)11-3311-06

DOI:10.11772/j.issn.1001-9081.2017.11.3311

基于时间序列分析和卡尔曼滤波的霾预报技术

张恒德¹, 咸云浩^{2*}, 谢永华^{2,3}, 杨乐², 张天航¹

(1. 中国气象局 国家气象中心, 北京 100081; 2. 南京信息工程大学 计算机与软件学院, 南京 210044;

3. 南京信息工程大学 江苏省网络监控中心, 南京 210044)

(*通信作者电子邮箱 yhaoxian_nuist@163.com)

摘要:为了提高霾预报的准确率,解决时序模型的预测延时和准确率不高的问题,提出了一种基于时间序列分析和卡尔曼滤波相结合的混合霾预报算法。首先,利用图检验法和单位根检验法(ADF)检验时间序列的平稳性,通过差分运算将非平稳序列转化成平稳序列,对转化后的平稳序列进行建模;然后,将得到的模型方程作为卡尔曼滤波的状态方程和观测方程,依靠卡尔曼滤波递推性进行预报。实验结果表明,采用时间序列分析和卡尔曼滤波相结合的混合霾客观预报订正方法能有效提高霾预测精度。

关键词:时间序列; 卡尔曼滤波; 能见度; 霾; 预报模型

中图分类号:TP391.9 **文献标志码:**A

Haze forecast based on time series analysis and Kalman filtering

ZHANG Hengde¹, XIAN Yunhao^{2*}, XIE Yonghua^{2,3}, YANG Le², ZHANG Tianhang¹

(1. National Meteorological Center, China Meteorological Administration, Beijing 100081, China;

2. School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China;

3. Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information and Technology, Nanjing Jiangsu 210044, China)

Abstract: In order to improve the accuracy of haze forecast and resolve the time lagging and low accuracy of temporal model, a mixed forecast method based on time series analysis and Karman filter was proposed. Firstly, the stability of time series was tested by graph analysis and eigenvalue analysis (ADF). Unstable time series were converted to stable ones by differential operation. A statistical function was established based on the stable time series. And then, the obtained model equations were used as the state and observation equation for Kalman filtering. Final haze forecast was based on recursion by Karman filtering. The experimental results showed that the accuracy of haze forecast is effectively improved by the mixed forecast method based on time series analysis and Karman filtering.

Key words: time series; Kalman filtering; visibility; haze; forecasting model

0 引言

霾是悬浮在大气中的颗粒物积累到一定程度,使水平能见度小于10 km 的一种天气现象^[1]。近年来,中国中东部霾事件频发,霾过程中低能见度会引发交通事故,高气溶胶浓度会影响空气质量及气候,也会引发多种呼吸道疾病,危害人类健康,受到各级政府、部门及社会各界的广泛关注。因此,急需建立和发展高精度的霾预报系统,进而加强霾的防御和防治。

目前霾预报主要有天气学预报法、数值预报法和统计学预报法。天气学预报法通过分析和诊断霾形成的天气学条件,结合预报员的经验和实况,用外推方法预报霾过程。廖碧婷等^[2]利用1998—2008年观测和模式资料计算了垂直交换系数,并利用垂直交换系数对典型霾过程进行预报。但天气学预报法的预报水平和精细化程度远远满足不了用户对预报服务的需求。

数值预报法主要通过空气质量数值预报模式预报能见度和主要污染物浓度,从而预报霾过程。中国气象局自主研发

的雾霾数值预报系统(CUACE/haze-fog)^[3]对主要雾霾过程预报效果较好,但对过程中污染物峰值浓度及能见度预报存在一定偏差。珠江三角洲空气质量数值预报系统能基本预报出霾过程中污染物的产生、聚集和消亡过程和时间节点,但低估了污染物的浓度^[4]。Wang等^[5]建立的长三角空气质量预报系统预报的上海和南京空气污染指数的准确率分别为50%~83%和80%。北京区域环境气象数值预报系统(Beijing Regional Environmental Meteorology Prediction System, BREMPS)对2014年京津冀PM2.5浓度的预报效果较好,相关系数在0.6以上,但预报值总体低于观测值,且24 h之后预报效果略有下降^[6]。由于现阶段模式中污染物排放、传输和沉降过程中的各参数均存在较大不确定性,导致预报的污染物浓度也存在较大不确定性^[5]。

统计学预报法是通过分析霾天气出现时的大尺度天气系统活动规律,建立中低空大气温度、相对湿度和风速等气象要素和能见度、污染物浓度的统计关系,最终得到霾的统计预报模型。已有研究利用最小二乘法、回归分析、时间序列分析、

收稿日期:2017-05-22;修回日期:2017-08-16。 基金项目:国家重点研发计划项目(2016YFC0203301)。

作者简介:张恒德(1977—),男,安徽含山人,高级工程师,博士,主要研究方向:环境气象、灾害性天气; 咸云浩(1991—),男,江苏淮安人,硕士研究生,主要研究方向:人工智能、机器学习、污染物浓度预报; 谢永华(1976—),男,江苏靖江人,教授,博士,主要研究方向:人工智能、图像处理; 杨乐(1990—),男,江苏淮安人,硕士研究生,主要研究方向:人工智能、机器学习、污染物浓度预报; 张天航(1987—),男,湖北荆门人,工程师,博士,主要研究方向:数值环境气象预报。



遗传算法、神经网络、卡尔曼滤波法、小波分析、支持向量机等方法预报霾。邹乐强^[7]基于气象要素、天气形势、能见度以及污染物浓度资料,建立了1~3 d 霾集成预报系统,对有无霾预报准确率较高,但对重度霾的预报准确率偏低。高辉^[8]基于K最近邻(K Nearest Neighbor, KNN)数据挖掘算法构建了北京地区霾等级预报模型,对轻度霾、中度霾和重度霾的空报率仅为4.7%、1.4% 和 2.6%。这些预测方法各有各的特点和适用场合,但由于污染物浓度受排放量、气象条件、谱分布以及化学成分和性质等的共同影响,用单一统计方法建立准确度较高的统计预报模型难度较大。将各预报模型进行合理的组合,可以使各个模型的特点相互补充,达到更加准确的预报效果^[9]。已有文献利用时间序列分析结合卡尔曼滤波应用在导航算法^[10]、交通路段行程预测^[11]和微机械陀螺仪漂移误差补偿^[12]中,尚没有研究将时间序列分析结合卡尔曼滤波法应用至霾预报中。因此,本文选择时间序列分析和卡尔曼滤波进行混合建模,有利于降低预报误差。

综上所述,基于霾预报的重要性和难度,本文将采用时间序列分析方法,建立能见度预报模型,再结合卡尔曼滤波订正方法建立能见度预报订正方程,并排除降水、沙尘暴、扬沙、浮尘、吹雪、雪暴等天气现象,进而得到霾预报结果。本文旨在建立具有更高实时性和精度的霾预报模型,这将有助于提高霾预报能力,为政府防控提供科学参考。

1 基于时间序列和卡尔曼滤波霾预报模型

1.1 时间序列分析

Box 等^[13]在20世纪70年代初提出了差分自回归滑动平均(AutoRegressive Integrated Moving Average, ARIMA)模型。它是指在将非平稳化的时间序列经过差分运算转化为平稳的时间序列,然后将预报量仅对它的滞后值以及随机误差的现值和滞后值进行差分自回归滑动平均的建模。其基本思想是用一定的数学模型来描述预报量随时间变化而形成的一组随机有序时间序列,通过时间序列建立的模型结合预报量的过去值和现在值预报未来值。ARIMA(p, d, q)模型由AR(p)模型、平稳化时间序列所做的差分次数 d 和MA(q)模型组成。AR(p)模型为自回归模型,其中 p 为自回归项;差分次数 d 是将时间序列由非平稳序时间序列通过差分运算转化成平稳时间序列所需的次数;MA(q)为滑动平均模型,其中 q 为滑动平均项数。

ARIMA模型建模的基本过程如下:

- 1) 分别求取该样本序列的自相关系数(Auto Correlation Function, ACF)和偏相关函数(Partial Auto Correlation Function, PACF),通过图检验法初步判断样本序列的平稳性。然后通过单位根检验法ADF来验证序列的平稳性。
- 2) 通过图检验法和ADF检验后,如果样本序列是非平稳的,那么需要对非平稳样本序列进行差分运算,直到非平稳序列转化称平稳序列为为止。
- 3) 对经过平稳化后的样本序列使用贝叶斯信息量(Bayesian Information Criterion, BIC)对模型定阶。
- 4) 模型确定后,对模型参数进行估计,并且检验模型是否具有统计意义。
- 5) 检验残差序列是否为白噪声,进行假设性检验。
- 6) 使用通过检验的模型进行预报。

1.2 卡尔曼滤波的原理和应用

卡尔曼滤波方法采用量测方程和状态方程组成的线性随机系统的状态空间模型来描述滤波器,并且利用状态方程的递推性,按线性无偏最小均方差估计准则,采用递推算法对滤波器的状态变量作出最佳估计,从而求得滤掉噪声后有用信号的最佳估计。大气环境预报领域中的卡尔曼滤波基本公式如下:

$$\mathbf{Y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{v}_t \quad (1)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\Phi}_{t-1} \boldsymbol{\beta}_{t-1} + \mathbf{w}_{t-1} \quad (2)$$

式(1)称为量测方程,式(2)称为状态方程。式中, $\boldsymbol{\beta}_t = [\beta_1, \beta_2, \dots, \beta_m]^T_t$ 为卡尔曼滤波系统状态向量,也称回归系数; $\boldsymbol{\Phi}$ 为状态转移矩阵; \mathbf{Y}_t 为测量向量即预报量, $\mathbf{Y}_t = [Y_1, Y_2, \dots, Y_m]^T_t$; \mathbf{X}_t 为预报因子矩阵。假定 $\boldsymbol{\Phi}$ 为单位矩阵, 经过简化后的卡尔曼滤波状态方程为:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{v}_{t-1} \quad (3)$$

假定 \mathbf{w} 和 \mathbf{v} 是两个互不相关的随机向量,应用最小二乘法,即可得到卡尔曼滤波的递推方程组,即:

$$\hat{\mathbf{Y}}_t = \mathbf{X}_t \hat{\boldsymbol{\beta}}_{t-1} \quad (4)$$

$$\mathbf{R}_t = \mathbf{C}_{t-1} + \mathbf{W} \quad (5)$$

$$\boldsymbol{\sigma}_t = \mathbf{X}_t \mathbf{R}_t \mathbf{X}_t^T + \mathbf{V} \quad (6)$$

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{X}_t^T \boldsymbol{\sigma}_t^{-1} \quad (7)$$

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{t-1} + \mathbf{A}_t (\mathbf{Y}_t - \hat{\mathbf{Y}}_t) \quad (8)$$

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \boldsymbol{\sigma}_t \mathbf{A}_t^T \quad (9)$$

式(4)~(9)组成的递推系统中: \mathbf{W} 和 \mathbf{V} 分别表示动态噪声和量测噪声的方差阵, $\hat{\mathbf{Y}}_t$ 为 t 时刻的预报值, \mathbf{X}_t 为预报因子, $\hat{\boldsymbol{\beta}}_{t-1}$ 为回归系数估计值, \mathbf{R}_t 为 $\hat{\boldsymbol{\beta}}_{t-1}$ 外推值的误差方差阵, \mathbf{C}_{t-1} 为 $\hat{\boldsymbol{\beta}}_{t-1}$ 滤波值的误差方差阵, $\boldsymbol{\sigma}_t$ 是预报误差方差阵, \mathbf{A}_t 为增益矩阵, \mathbf{Y}_t 为 t 时刻的观测值即能见度实况值。

卡尔曼滤波订正方法通过不断递推,不断更新实测信息,采用更新实测的数据对状态估计进行修正,从而对预报产品进行订正,提高预报准确率。

1.3 时序差分自回归滑动平均与卡尔曼滤波模型

本方案中采用的资料包括:气象观测资料、欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF)细网格数值预报产品。在实验之前,采用双线性插值法,将ECMWF格点数据转化成站点数据作为实验数据。

本方案首先选用时序差分自回归滑动平均模型得到最优的能见度预报模型,然后结合使用卡尔曼滤波订正方法建立能见度客观预报订正模型,再排除降水、沙尘暴、扬沙、浮尘、吹雪、雪暴等视程障碍天气,得到霾预报结果。模型流程如图1所示,具体步骤如下:

- 1) 选取4个典型代表站(北京、南京、广州、上海),根据对霾天气的分析,利用中国气象局实况资料,构建时序差分自回归滑动平均模型。
- 2) 将各代表站观测值作为样本序列,选取700个数据进行建模,使用图检验法和根检验法判断序列是否平稳,如果样本序列为非平稳序列,则对非平稳序列进行差分运算,直至样本序列平稳为止。
- 3) 分别计算经过平稳化的各代表站样本序列的自相关



函数(ACF)和偏自相关函数(PACF)。根据 BIC 定阶并确定 ARIMA 模型, 然后检查模型的有效性。确定混合模型流程(如图 1 所示)。

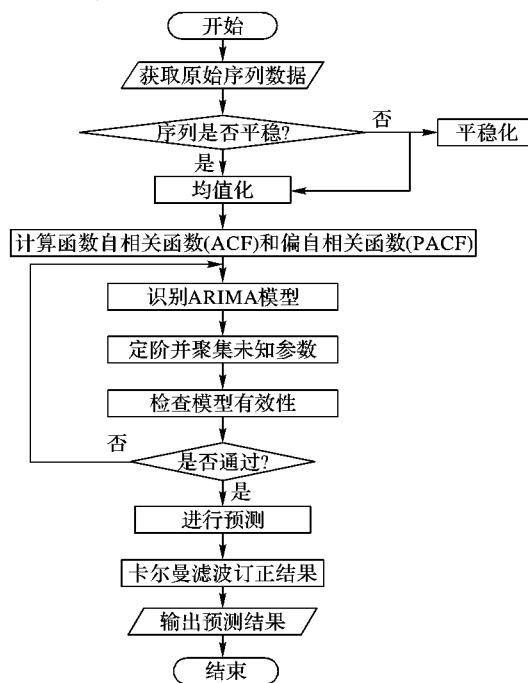


Fig. 1 Model process

4) 利用 ECMWF 模式产品, 将 ARIMA 模型确定的方程模型作为卡尔曼滤波的初始量测方程, 然后根据 ECMWF 模式产品确定 \mathbf{W}_0 、 \mathbf{V}_0 、 \mathbf{C}_0 、 $\boldsymbol{\beta}_0$ 。由于本文使用时间序列分析建立能见度最优预报方程, 所以将时间序列分析预报方程中的回归系数作为卡尔曼滤波过程中回归系数的初始值。由于 $\boldsymbol{\beta}_0$ 是从样本资料中通过严格的数学计算得到的, 因此理论上认为它是十分精确的, 即与理论值相等, 所以 \mathbf{C}_0 是 m 阶零方阵,

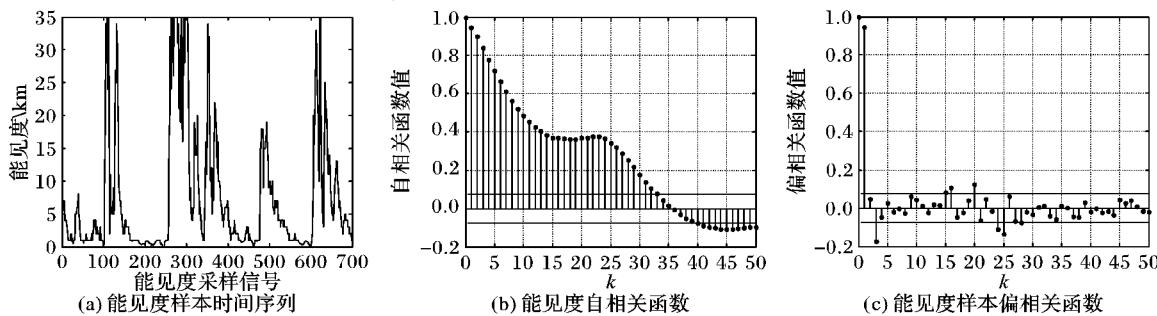


Fig. 2 Visibility experimental results of Beijing station

根据 BIC 确定 ARIMA 模型, 最小的 BIC 值为 3.561×10^3 , 所以 $p = 1, q = 2$ 。至此 p, q 定阶完成。由此北京站能见度样本序列的时间序列模型为 ARIMA(1, 0, 2) 模型。由表 1 模型参数计算获得该模型方程:

$$X(t) = 0.933345X(t-1) + \alpha_t + 0.0330181\alpha_{t-1} - 0.191094\alpha_{t-2} \quad (11)$$

式中: α_t 为残差; $0.0330181\alpha_{t-1}$ 对预报方程影响较少, 可以省去, 所以公式如式(12):

$$X(t) = 0.933345X(t-1) + \alpha_t - 0.191094\alpha_{t-2} \quad (12)$$

同理, 选取南京站连续 700 h 的数据作为能见度样本时

即 $\mathbf{C}_0 = [0]_{m \times m}$ 。经过实验证明状态量测噪声 \mathbf{V}_0 和输入噪声 \mathbf{W}_0 的初值对回归系数 $\boldsymbol{\beta}$ 的订正并不敏感, 因此, 本文将 $\mathbf{W}_0, \mathbf{V}_0$ 都设为零方阵。最后根据卡尔曼滤波递推公式, 对时间序列建立的能见度预报方程动态修改权重。将卡尔曼滤波订正能见度预报方程的过程及误差订正过程设计成自动批处理程序。

5) 根据霾天气发生时能见度、相对湿度、降水等气象要素进一步判断, 得到霾预报结果。最后利用观测实况进行预报检验。选用气象上衡量预报准确率的校正得分率(True Shooting percentage, TS) 评分算法, 对基于时间序列分析和卡尔曼滤波的霾预报结果进行评分。TS 评分的定义如下:

$$TS = \frac{t_1}{t_1 + t_2 + t_3} \times 100\% \quad (10)$$

其中: t_1 为正确预报霾的次数, t_2 为漏报霾的次数, t_3 为空报霾的次数。由定义看出, TS 越大, 预报效果就越好, 预报准确率就越高。

2 实验结果

2.1 实验资料和平台

本文选取北京、南京、杭州和广州 4 个典型站点的观测值作为实验数据。将观测值(逐小时能见度实况资料)作为样本序列 $\{Z_t\}$, 运用时间序列分析分别对 4 个典型站点进行建模。

选取北京站连续 700 h 的数据作为能见度样本时间序列 $\{Z_t\}$ (如图 2(a)所示), 然后求取北京站样本时间序列的前 50 个自相关系数和偏相关函数(图 2(b)、(c))。北京站能见度样本资料自相关函数大致按指数衰减至零, 可以初步推断原始能见度时间序列是平稳序列。同时本文使用单位根检验法(ADF 检验)来检验北京站能见度时间序列, 原始序列平稳性检验结果: $h = 1$ (这里的 h 为平稳性检验标志, 非零代表平稳, 零代表非平稳)。检验结果与之前通过分析自相关系数图得出的结论是一致的, 即北京站能见度样本时间序列是平稳数据, 可以用来作为预报序列。

间序列 $\{Z_t\}$ (如图 3(a)所示), 求取的前 50 个自相关系数和偏相关函数(图 3(b)、(c))。南京站能见度样本资料自相关函数也很快衰减至零, 初步推断原始能见度时间序列是平稳序列, 进一步使用单位根检验法(ADF 检验)检验出南京站能见度时间序列是平稳数据, 与之前通过分析自相关系数图得出的结论是一致的, 可以作为预报序列。

根据 BIC 确定 ARIMA 模型, 最小的 BIC 值为 2.6614×10^3 , 所以 $p = 5, q = 4$ 。至此 p, q 定阶完成。由此南京站能见度样本序列的时间序列模型为 ARIMA(5, 0, 4) 模型。由表 2 模型参数计算获得该模型方程为式(13):



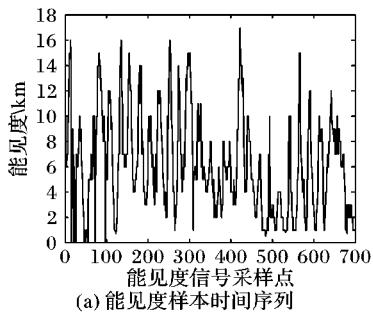
$$\begin{aligned} X(t) = & 1.51105X(t-1) + 0.388007X(t-2) - \\ & 0.92432X(t-3) - 0.596349X(t-4) + \\ & 0.606433X(t-5) + \alpha_t + 0.661273\alpha_{t-1} + \\ & 0.872461\alpha_{t-2} - 0.228326\alpha_{t-3} - 0.475871\alpha_{t-4} \end{aligned} \quad (13)$$

表1 北京站 ARIMA 模型参数表

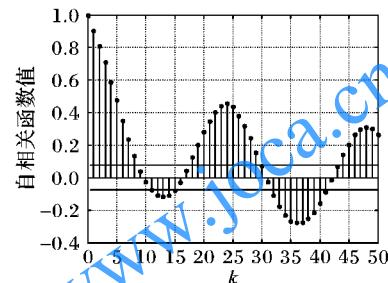
Tab. 1 ARIMA model parameter table of Beijing station

参数	参数值	标准差	T统计
Constant	0.4939620	0.3594950	1.37405
AR{1}	0.9333450	0.0156674	59.57230
MA{1}	-0.03030181	0.0161219	-2.04802
MA{2}	0.1910940	0.0195544	9.77247
Variance	9.1324700	0.2580780	35.38640

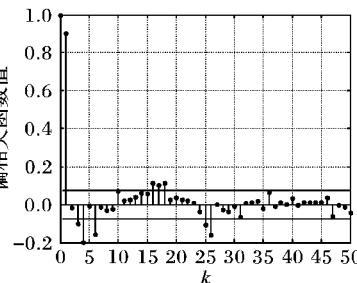
同理,选取杭州站连续700个小时的数据作为能见度样本时间序列 $\{Z_t\}$,如图4(a)所示,求取的前50个自相关系数和偏相关函数(图4(b)、(c))。杭州站能见度样本资料自相关函数也很快衰减至零,初步推断原始能见度时间序列是平稳序列,进一步使用单位根检验法(ADF检验)检验出南京站能见度时间序列是平稳数据,可以作为预报序列。根据BIC



(a) 能见度样本时间序列



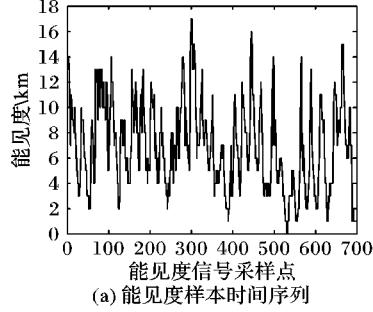
(b) 能见度样本自相关函数



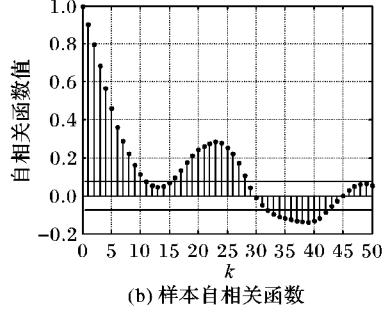
(c) 能见度样本偏相关函数

图3 南京站能见度实验结果

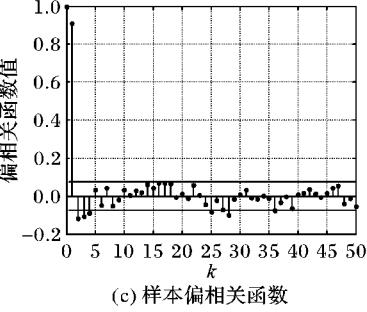
Fig. 3 Visibility experimental results of Nanjing station



(a) 能见度样本时间序列



(b) 样本自相关函数



(c) 样本偏相关函数

图4 杭州站能见度实验结果

Fig. 4 Visibility experimental results of Hangzhou station

同理,选取广州站连续700 h的数据作为样本时间序列 $\{Z_t\}$,如图5(a)所示,求取的前50个自相关系数和偏相关函数(图5(b)、(c))。广州站能见度样本资料自相关函数也很快衰减至零,初步推断原始能见度时间序列是平稳序列,进一步使用单位根检验法(ADF检验)检验出广州站能见度时间序列是平稳数据,可以作为预报序列。根据BIC确定ARIMA模型,最小的BIC值为 3.562×10^3 ,所以 $p=6, q=5$ 。至此 p, q 定阶完成。由此广州站能见度样本序列的时间序列模型为ARIMA(6,0,5)模型。

由表4模型参数计算获得该模型方程为式(15):

$$\begin{aligned} X(t) = & 0.91651X(t-1) + 1.00275X(t-2) - \\ & 0.29851X(t-3) - 0.756201X(t-4) - \end{aligned}$$

准则确定ARIMA模型,最小的BIC值为 2.4967×10^3 ,所以 $p=2, q=1$ 。所以至此 p, q 定阶完成。由此杭州站能见度样本序列的时间序列模型为ARIMA(2,0,1)模型。由表3模型参数计算获得该模型方程为式(14):

$$\begin{aligned} X(t) = & 1.62102X(t-1) - 0.670331X(t-2) + \alpha_t + \\ & 0.607047\alpha_{t-1} \end{aligned} \quad (14)$$

表2 南京站 ARIMA 模型参数表

Tab. 2 ARIMA model parameter table of Nanjing station

参数	参数值	标准差	T统计
Constant	0.0967424	0.0324409	2.98212
AR{1}	1.5110500	0.1055890	14.31060
AR{2}	0.3880070	0.2035850	1.90587
AR{3}	-0.9243200	0.1736280	-5.32357
AR{4}	-0.5963490	0.2179340	-2.73637
AR{5}	0.6064330	0.0899210	6.74407
MA{1}	-0.6612730	0.1105740	-5.98035
MA{2}	-0.8724610	0.1332490	-6.54758
MA{3}	0.2283260	0.1518960	1.50317
MA{4}	0.4758710	0.1173810	4.05407
Variance	2.3881200	0.0582432	41.00260

$$\begin{aligned} 0.632804X(t-5) + & 0.738651X(t-6) + \\ & \alpha_t + 0.432253\alpha_{t-1} + 1.25093\alpha_{t-2} - 0.404755\alpha_{t-3} - \\ & 0.356343\alpha_{t-4} - 0.123638\alpha_{t-5} \end{aligned} \quad (15)$$

表3 杭州站 ARIMA 模型参数表

Tab. 3 ARIMA model parameter table of Hangzhou station

参数	参数值	标准差	T统计
Constant	0.339427	0.0938038	3.61848
AR{1}	1.621020	0.0991168	16.35470
AR{2}	-0.670331	0.0896005	-7.48133
MA{1}	-0.607047	0.1116620	-5.43648
Variance	1.996490	0.0719921	27.73210



随后将以上通过时序模型得到的北京站、南京站、杭州站和广州站预报方程用于卡尔曼滤波,通过卡尔曼滤波订正时序模型求得结果。

表 4 广州站 ARIMA 模型参数表

Tab. 4 ARIMA model parameter table of Guangzhou station

参数	参数值	标准差	T 统计
Constant	0.295374	0.0586255	5.03832
AR{1}	0.916510	0.0360402	25.43020
AR{2}	1.002750	0.0518295	19.34700
AR{3}	-0.298510	0.0581609	-5.13249
AR{4}	-0.756201	0.0555799	-13.60570
AR{5}	-0.632804	0.0504084	-12.55350
AR{6}	0.738651	0.0282139	26.18040
MA{1}	-0.432253	0.0410538	-10.52890
MA{2}	-1.250930	0.0468730	-26.68760
MA{3}	0.404755	0.0655493	6.17482
MA{4}	0.356343	0.0484183	7.35967
MA{5}	0.123638	0.0439512	2.81307
Variance	8.486100	0.3808260	22.28340

2.2 实验结果分析

2.2.1 本方法实验结果

在确定了时间序列差分自回归滑动平均和卡尔曼滤波的

霾客观预报订正模型后,做北京、南京、杭州、广州预报实验。实验日期:2014 年 10 月—2015 年 2 月。实验结果如图 6 所示。结果显示,北京、南京、杭州和广州 4 个站点的实验数据散点图中能见度客观预报订正结果与实况的波形图基本上一致,说明时间序列分析和卡尔曼滤波的霾客观预报模型能很好地反映实况能见度的变化趋势。将时间序列差分自回归滑动平均建立的能见度预报结果与经历过卡尔曼滤波订正之后的结果进行对比发现,经过卡尔曼滤波订正后的结果与能见度实况拟合情况更好。在实验过程中发现,如果只用能见度预报霾是否发生,则误差会很大,所以将能见度、风速、相对湿度和温度露点差结合建立霾的预报模型。结果显示:利用 ECMWF 模式预报因子数据和霾预报模型对 2014 年—2015 年 2 月进行预报实验来检验基于时间序列差分自回归滑动平均和卡尔曼滤波的霾预报模型的预报准确率。预报准确率较 CUACE 有明显提高。

2.2.2 与数值预报模式结果对比分析

利用本文的模型输出结果与中国气象局第一代雾霾数值预报业务系统(CUACE/haze-fog_54km)输出结果进行对比分析。选取 2013 年 10 月—2014 年 2 月各代表站点的霾天气进行预报实验,利用每日 8:00 的资料进行对比检验,检验结果见表 5。

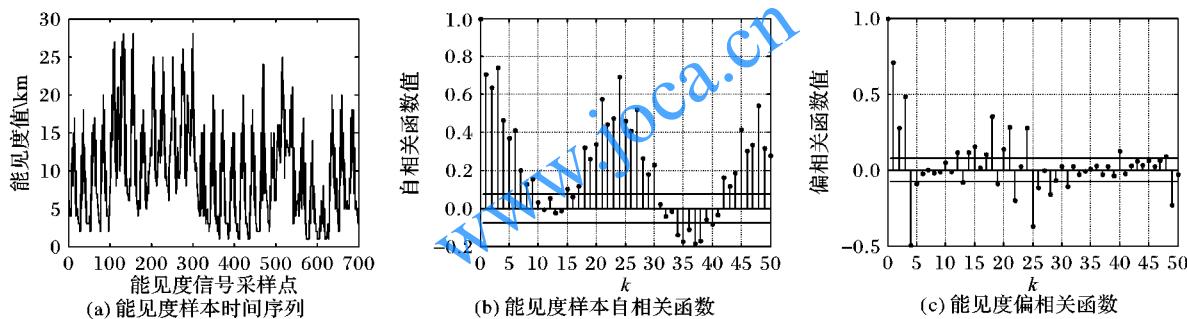


Fig. 5 Results of visibility experiment of Guangzhou station

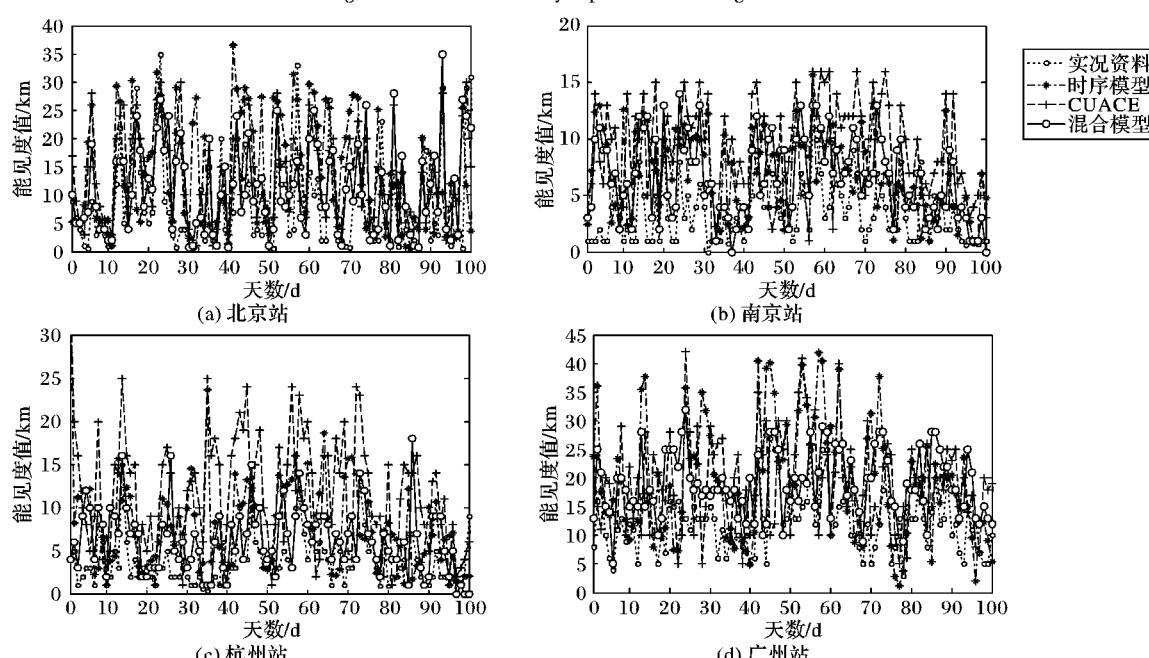


Fig. 6 Comparison of visibility forecast in Beijing, Nanjing, Hangzhou and Guangzhou



由检验的对比结果可看出,利用时间序列差分自回归滑动平均与卡尔曼滤波建立的预报模型得到的霾预报结果与CUACE模式在北京、南京、杭州、广州的准确率相比分别提高了11%、13%、11%、1% (平均提高了10个百分点),而空报率和漏报率则有不同程度减低,平均值分别降低了3个百分点和7个百分点。进一步验证了本文的方法在霾预报上有着更好的效果。

表5 本文和 CUACE 模式预报结果对比

Tab. 5 Compared with the CUACE model prediction results

城市	空报率/%		漏报率/%		TS 评分	
	CUACE	本文方法	CUACE	本文方法	CUACE	本文方法
北京	31	25	8	3	0.61	0.72
南京	36	27	10	6	0.54	0.67
杭州	34	29	9	3	0.57	0.68
广州	10	19	20	10	0.70	0.71
平均	28	25	12	5	0.60	0.70

3 结语

为提高霾预报准确率,本文构建了时间序列差分自回归滑动平均和卡尔曼滤波混合的霾客观预报订正模型:先通过时间序列分析推导出能见度预报方程,然后利用卡尔曼滤波递推性动态订正预报过程中的权重,避免了时间序列建立低阶模型预报准确率低的情形,也降低了卡尔曼滤波在建立初始状态方程和测量方程的难度。该混合算法与 CUACE 模式模拟结果相比,能有效提高霾预报准确率。

目前本文研究内容仅限于北京、南京、杭州、广州几个典型站点,且客观预报准确率还有待提高,后续研究将增加实验站点、延长实验时间序列、改进算法,进一步提高客观预报准确率。

参考文献 (References)

- [1] 吴兑. 关于霾与雾的区别和灰霾天气预警的讨论[J]. 气象, 2005, 31(4): 3 - 7. (WU D. Discussion about the differences between haze and fog and the gray smog weather warning[J]. Meteorological Monthly, 2005, 31(4): 3 - 7.)
- [2] 廖碧婷, 吴兑, 陈静, 等. 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, 2012, 28(3): 417 - 424. (LIAO B T, WU D, CHEN J, et al. The forecast application of grey haze weather variations and vertical exchange coefficients [J]. Journal of Tropical Meteorology, 2012, 28(3): 417 - 424.)
- [3] ZHOU C H, GONG S, ZHANG X Y, et al. Towards the improvements of simulating the chemical and optical properties of Chinese aerosols using an online coupled model: CUACE/Aero[J]. Tellus Series B: Chemical & Physical Meteorology, 2012, 64(1): 91 - 102.
- [4] 邓雪娇, 邓涛, 吴兑, 等. 珠江三角洲空气质量与能见度数值预报模式系统[J]. 广东气象, 2010, 32(4): 18 - 22. (DENG X J, DENG T, WU D, et al. The numerical forecast model system of air quality and visibility over the Pearl River Delta region [J]. Guangdong Meteorological, 2010, 32(4): 18 - 22.)
- [5] WANG T J, JIANG F, DENG J J, et al. Urban air quality and regional haze weather forecast for Yangtze River Delta region[J]. Atmospheric Environment, 2012, 58(15): 70 - 83.
- [6] 赵秀娟, 徐敬, 张自银, 等. 北京区域环境气象数值预报系统及PM2.5 预报检验[J]. 应用气象学报, 2016, 27(2): 160 - 172. (ZHAO X J, XU J, ZHANG Z Y, et al. Beijing regional environmental meteorological numerical forecasting system and PM2.5 forecast test[J]. Journal of Applied Meteorology, 2016, 27(2): 160 - 172.)
- [7] 邹乐强. 最小二乘法原理及其简单应用[J]. 科技信息, 2010, 2 (23): 282 - 283. (ZOU L Q. The principle of the least square method and its simple application [J]. Technology Information, 2010, 2(23): 282 - 283.)
- [8] 高辉. 几类常用非线性回归分析中最优模型的构建与 SAS 智能化实现[D]. 北京: 中国人民解放军军事医学科学院, 2012. (GAO H. The construction of optimal model in several types of non-linear regression analysis and the intelligent realization of SAS[D]. Beijing: The PLA Academy of Military Medicine, 2012.)
- [9] 原继东, 王志海. 时间序列的表示与分类算法综述[J]. 计算机科学, 2015, 43(3): 1 - 7. (YUAN J D, WANG Z H. The representation and classification algorithm of the time series [J]. Computer Science, 2015, 43(3): 1 - 7.)
- [10] 周俊, 张鹏, 刘成良. 基于时间序列分析的卡尔曼滤波组合导航算法[J]. 农业工程学报, 2010, 26(12): 254 - 258. (ZHOU J, ZHANG P, LIU C L. Based on time series analysis of integrated navigation Kalman filtering algorithm [J]. Journal of Agricultural Engineering, 2010, 26(12): 254 - 258.)
- [11] 田甜, 王秀玲, 吕芳. 基于 Kalman 和 ARIMA 组合模型的路段行程时间预测[J]. 信息技术, 2016(3): 148 - 150, 155. (TIAN T, WANG X L, LYU F. Urban travel time prediction based on Kalman and ARIMA model[J]. Information Technology, 2016(3): 148 - 150, 155.)
- [12] 李杰, 张文栋, 刘俊. 基于时间序列分析的 Kalman 滤波方法在 MEMS 陀螺仪随机漂移误差补偿中的应用研究[J]. 传感技术学报, 2006, 19(5): 2215 - 2219. (LI J, ZHANG W D, LIU J. Kalman filtering method based on time series analysis in the application of MEMS gyroscope random drift error compensation study [J]. Journal of Sensing Technology, 2006, 19(5): 2215 - 2219.)
- [13] BOX G E P, JENKINS G M. Time Series Analysis: Forecasting and Control[M]. San Francisco: Holden Day, 1976: 303.
- [14] 周国强, 韩晟. 基于 ARIMA 模型的中国医院药品费用增长趋势分析[J]. 药品评价, 2014(14): 15 - 19. (ZHOU G Q, HAN S. Predictive analysis of China's hospital drug expenditure based on ARIMA model[J]. Drug Evaluation, 2014(14): 15 - 19.)

The work is partially supported by the National Key R&D Program (2016YFC0203301).

ZHANG Hengde, born in 1977, Ph. D., senior engineer. His research interests include environmental meteorology, catastrophic weather.

XIAN Yunhao, born in 1991, M. S. candidate. His research interests include artificial intelligence, machine learning, pollutant concentration forecast.

XIE Yonghua, born in 1976, Ph. D., professor. His research interests include artificial intelligence, image processing.

YANG Le, born in 1990, M. S. candidate. His research interests include artificial intelligence, machine learning, pollutant concentration forecast.

ZHANG Tianhang, born in 1987, Ph. D., engineer. His research interest include numerical environmental meteorological prediction.