

文章编号: 2096-1618(2022)06-0642-09

基于广播式自动相关监视系统的高空风场 短时预测方法研究

甄福东¹, 周淑玥², 陈丽晶¹, 朱嘉慧², 王海江²

(1. 中国民用航空西北地区空中交通管理局甘肃分局, 甘肃 兰州 730087; 2. 成都信息工程大学, 四川 成都 610225)

摘要:使用飞机作为天气传感器是空中交通管理和气象研究的最新进展,其数据有利于提高气象信息监测和预报的准确性,这将是航空气象领域的重大突破。结合 ADS-B 和 Mode S 数据的风场反演结果,更加丰富的数据集使短时风场的预测效果优于传统方法。利用一种新的、多功能、高精度的优化风场数据进行短时高空风场的预测研究。为提高中尺度高空风场短时预报的精度,基于高精度、高覆盖的风场反演数据,研究了高空风场的短时预测算法。通过分析高空风场时间序列的分布特征和变化趋势,建立 SARIMA 和 GPR 两种时间序列预测模型,分别讨论两种模型的优缺点及其应用方向,并对模型的预测准确性进行定量分析。结果表明,基于 GPR 的方法优于基于 SARIMA 的方法。在可接受的预报时间范围内,基于 GPR 的预测模型能较好地捕捉高空风场的变化趋势,具有较好的泛化能力。在可用数据较少的情况下,GPR 模型的复杂度不太高,时间相对较少,即用较少的时间换取更高的准确性,非常适合高空风场的短时预测。

关键词:广播式自动相关监视;二次雷达;SARIMA;GPR

中图分类号:TP311.13

文献标志码:A

doi:10.16836/j.cnki.jcuit.2022.06.005

0 引言

航线上的天气状况是影响航班运行和调度的主要因素,由于缺乏准确的高空风场测量信息,研究人员不得不依赖数值天气预报模型,该类模型通常会以较大的局部精度进行平滑处理从而保证较大区域上的准确度。与其他气象观测数据相比,高分辨率高空风场观测资料尤其稀缺,不能作为空中交通管制参考资料的^[1-2]。随着对高空风场资料的需求,各国发展二次雷达(secondary surveillance radar, SSR)和广播式自动相关监视系统(automatic dependent surveillance-broadcast, ADS-B)技术,它们不仅对航空交通管制做出了贡献,对解决气象领域数据匮乏的问题具有重要意义,而且在实际气象监测和航空作战中具有实用价值。因此,本文在先前的研究基础上,利用 ADS-B 反演风场结果进行短时风场的预测研究^[3]。

线性回归模型、季节性差分自回归滑动平均模型(seasonal autoregressive integrated moving average, SARIMA)和高斯过程回归模型(Gaussian process regression, GPR)多用于时间序列预测。第一种模型太简单,无法对时间序列中的复杂模式进行建模。SARIMA 模型效果较好,但存在均值收敛问题,这阻碍了其在长期时间序列预测中的性能,对长期预测问题建模

是不够的。高斯过程回归是一种非参数概率回归技术,由于其非参数特性,GPR 模型比 SARIMA 模型具有更灵活的建模能力^[4]。

GPR 模型是一种强大的机器学习算法,只依赖很少的参数进行预测。由于它几乎是非参数的,所以即使在可用数据很少的情况下,也能有效地解决各种监督学习问题^[5]。目前,基于机器学习的预测模型可以考虑时间相关的组件,如季节性、趋势、周期、不规则分量等,以最大化数据驱动预测和预测的准确性。本文研究了 SARIMA 模型和 GPR 模型的建模方法,用这两种模型预测高空风场的时间序列,并讨论了这两种模型在短时高空风场预测中的性能。

1 基于 SARIMA 模型的预测

1.1 ARIMA 模型

ARIMA 模型来源于自回归移动平均(ARMA)模型,该模型是自回归(AR)模型和移动平均(MA)模型的组合。结构表示为 $ARIMA(p, d, q)$, p 是自回归参数, d 是差分次数,即用于消除趋势和季节性并稳定时间序列的平均水平, q 是移动平均参数^[6]。模型由 AR 模型、差分模型和 MA 模型 3 部分组成,方程式如下:

$$AR(p): Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \quad (1)$$

$$MA(q): Y_t = \mu - \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

其中: Y 是时序观测值, μ 是该序列的平均值, α 和 β 分别是其对应的权重值, ε_t 是随机波动偏差, ε 是预测的残差。两个模型的混合即为ARMA(p,q)模型,从该模型出发对时间序列进行 d 次差分得到ARIMA(p,d,q)模型。

时间序列一般可以分解为长期趋势项、季节周期项和残差项。趋势项是指数据在很长一段时间内的上升和下降运动;季节周期项是数据具有季节性变化;残差项即噪声,是随机间隔的尖峰和低谷。预测平稳序列相对容易,预测更可靠,因为它消除了任何连续的自相关,因此预测模型中的预测变量几乎是独立的。时间序列的每个周期都有其独特的行为,如果序列不平稳,就无法预测或概括未来的时间段。因此,有必要在预测之前进行稳定检验,主要方法是利用原始数据的ADF(augmented dickey-fuller)检验,判断其单位根的存在性来定义平稳性^[7]。最初的假设是有一个单位根,只要统计值小于1%,原始假设就会被显著拒绝,数据被认为平稳。

时间序列的平稳方法一般采用差分法,通常从差分的最低值开始,即 $d=1$,使序列平稳。然而,如果仍然存在显著的趋势或自相关性,则滞后 ≥ 10 ,序列需要二阶差分。如果差分序列滞后1的自相关为零或负,或自相关较小且无规律,则序列不需要更多差分。在获得平稳序列后,需要选择模型和相应的参数。在此之前,需要了解自相关和偏自相关的概念。自相关衡量信号与其自身的相关性,而自相关函数图(autocorrelation function, ACF)是上述结果的图形表示。偏自相关是自相关的一个子集,偏自相关函数图(partial autocorrelation function, PACF)表示两个时间点观测值之间的相关性。ACF可以用来确定MA模型中使用的最佳参数,而AR模型的识别通常由PACF完成^[8]。

1.2 基于SARIMA模型的预测原理

虽然ARIMA模型可以处理带有趋势的数据,但不支持带有季节性成分的时间序列。实验中,由于时间序列的季节性,ARIMA的预测效果不佳。因此,本文将利用SARIMA模型对高空风场序列进行预测。该模型通过在ARIMA模型中加入额外的季节性项,添加3个新的超参数和额外的季节周期参数而形成。模型的季节性部分包括与模型的非季节性部分相似的项,但它们涉及季节周期的后移。SARIMA模型的建立过程如下:

- (1) 分析高空风场序列的趋势特征和季节性;
- (2) 时间序列数据平稳性判断,并对序列进行ADF检验;
- (3) 对于非平稳序列进行差分处理;

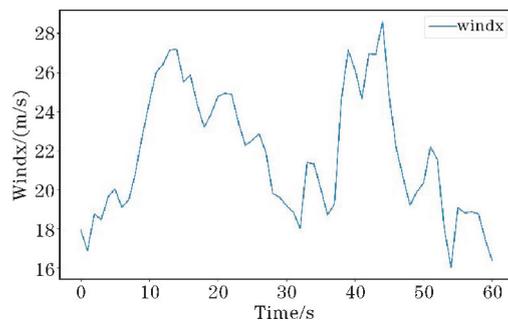
(4) 绘制相关图和自相关图;

(5) 根据上述结果构建SARIMA模型,确定相应的参数范围。

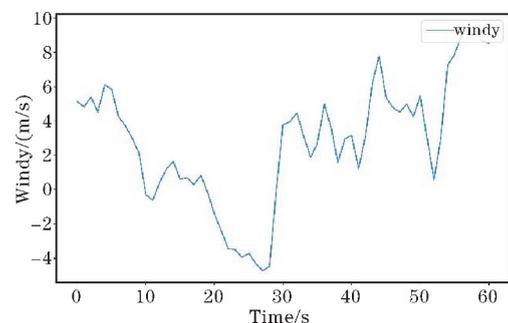
SARIMA模型有7个参数,其结构表示为SARIMA(p,d,q)(P,D,Q) m 。非季节参数 p,d,q 与ARIMA模型类似,而季节参数 P,D,Q 是参数 p,d,q 对应的季节参数, m 是单个季节性周期的时间步数。 m 会影响 P,D,Q 。在基于ARIMA模型配置SARIMA模型时,还需要为序列的季节性元素选择超参数。根据ACF和PACF的图表分布情况进一步确定参数信息,以明确最新时间步长的相关性。同样,也能够根据两个图表得出季节滞后时间步长的相关性,从而确定SARIMA的参数值^[9]。如果季节滞后 k 处的自相关为正,则将季节AR项(P)添加到模型中。若季节性滞后的自相关为负,则将季节性MA项(Q)添加到模型中。

1.3 风场时间序列分析和模型建立

在建立时间序列预测模型之前,本文分析了高空风场数据的时间序列。分析了季节性和平稳性,以获得未来预测的相关参数。以位置Point(0,0,11)为例对时间序列分析做说明,图1是高空风场 x 和 y 分量的原始序列图,显示了它们的变化趋势。首先,分析原始序列的季节性,分解数据并查看季节性。图2是相应的季节性分析图, x 和 y 分量的单个季节性周期的时间步长为9 min。从图中可以看出,该序列没有明显的趋势。然后进行ADF检验,发现高空风场原始数据的 x 和 y 分量非平稳,因此有必要对其进行一阶差分。



(a) x 分量



(b) y 分量

图1 高空风场原始序列图

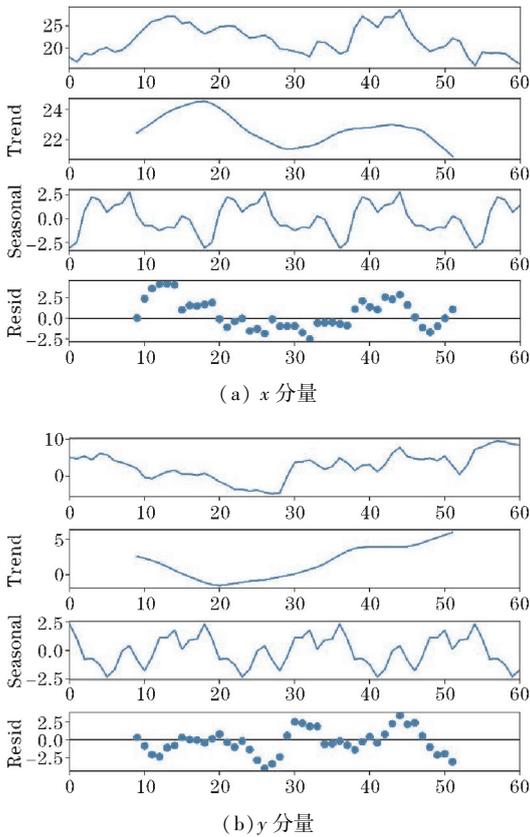


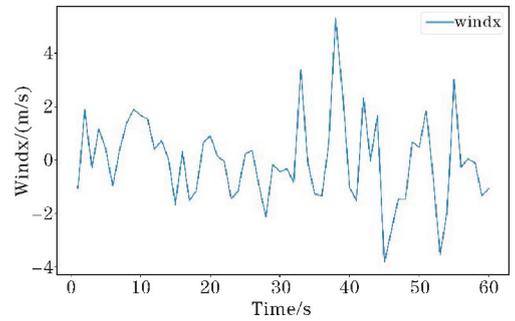
图2 高空风场原始序列季节性分析图

对非平稳数据进行差分处理,并对高空风场时间序列的 ADF 检验结果进行分析,该序列通过了一阶差分检验。取 $d=1$,并得到最终的差分序列图, x 和 y 分量的一阶差分序列如图 3 所示。同时,结合参数搜索法得到季节性预测和非季节性预测的相关参数,得到了合适的预测模型。本文计算了原始序列的 ACF 和 PACF,结果如图 4 所示,反映了时间序列中的观测值和前一时间步长中的观测值之间的关系强度。根据 ACF 和 PACF 图以及 ADF 检验结果判断序列是否平稳,ACF 函数图和 PACF 函数图显示了明显的截尾现象,再次验证了该序列是一个平稳时序。

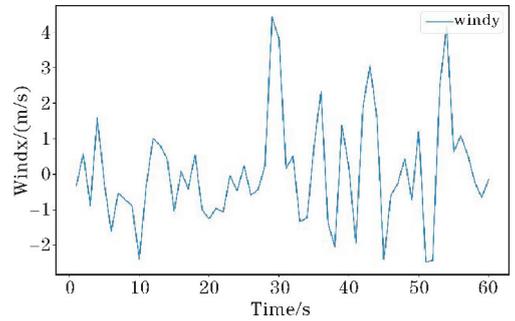
图 4 中,ACF 图显示了时间序列与其自身滞后的相关性,每条垂直线代表序列和从滞后零点开始的滞后之间的相关性,蓝色阴影区域是显著性水平,蓝色阴影区域之外的滞后是更大的滞后。PACF 显示任何给定时间序列滞后与当前序列之间的自相关,但消除了两者之间的滞后。从 ACF 图可以看出,除了第一条垂直线之外,几乎所有的滞后都在蓝色突出显示的区域,这表明不同滞后观测值之间的相关性几乎为零。最后,结合直接搜索结果,确定两个序列的模型如下:

$$\text{Model_windx}_{730} = \text{SARIMA}(p, d, q) (P, D, Q) m = \text{SARIMA}(1, 1, 1) (1, 1, 2) 18 \quad (3)$$

$$\text{Model_windy}_{730} = \text{SARIMA}(p, d, q) (P, D, Q) m = \text{SARIMA}(3, 1, 2) (2, 1, 0) 18 \quad (4)$$

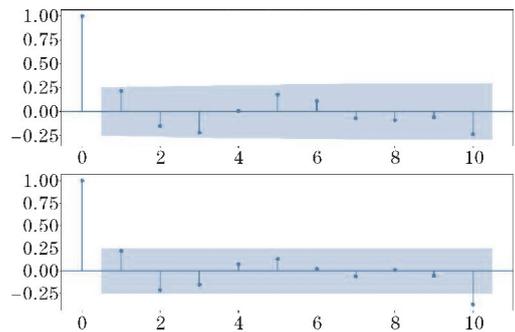


(a) x 分量

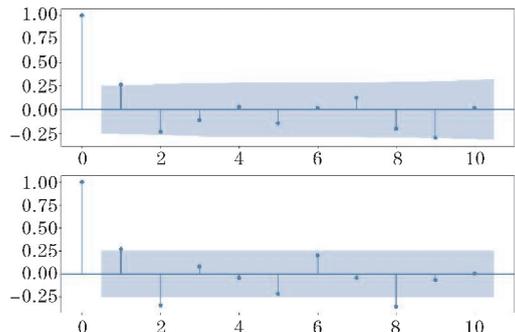


(b) y 分量

图3 高空风场一阶差分序列图



(a) x 分量



(b) y 分量

图4 高空风场一阶差分序列 ACF 和 PACF 图

2 基于高斯过程回归模型的预测

2.1 高斯过程模型

高斯过程的一种应用是通过监督学习进行回归,

被称为高斯过程回归(GPR)。GPR可以看作是一种机器学习算法,结合先验知识并找到一个函数来尽可能接近地描述一组给定的数据点。换句话说,高斯过程回归的主要思想是从无数函数中找到满足测试数据的任何函数集,根据一组数据集的先验信息不断缩小找到的函数集的范围,最后通过贝叶斯法则的方法和高斯分布的性质来计算找到的函数集的分布,并根据函数集的分布对后续数据进行预测^[10]。

高斯过程由一个过程 $f(x)$ 的均值函数 $m(x)$ 和核函数 $k(x, x')$ 确定,其中 x 和 x' 表示两个不同的样本输入随机变量。那么 $f(x)$ 的高斯过程即可表示为

$$m(x) = E[f(x)] \quad (5)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))^T] \quad (6)$$

$$f(x) \sim GP(m(x), k(x, x')) \quad (7)$$

2.2 核函数选择与超参数优化

核函数是高斯过程的核心,决定了高斯过程的性质。在高斯过程中生成一个协方差矩阵来衡量任意两点之间的距离。矩阵描述了分布的形状,决定了预测函数的特性。核函数的作用是在空间上平滑先验分布,即通过先验约束使相邻点具有更大的相似度,其测量值或估计值也具有更大的相似度,从而使拟合函数具有更好的空间平滑性。同时,生成的协方差矩阵不仅描述了分布的形状,还决定了预测函数的特性。因此,可以建立训练集数据 y 和测试集 y' 的联合分布^[11]:

$$\begin{bmatrix} y \\ y' \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(x, x) & K(x, x') \\ K(x', x) & K(x', x') \end{bmatrix}\right) \quad (8)$$

由此可以算出待测试数据 y' 的后验概率分布。

$$p(f' | x, x', y) \sim N(\mu', \Sigma') \quad (9)$$

不同的核函数有不同的衡量方法,得到的高斯过程的性质也不同。本文将讨论3个常见的核函数:平方指数核(squared exponential, SE)、有理二次核(rational quadratic kernel, RQ)和周期性内核(periodic kernel, PER)。

2.2.1 平方指数核(SE)

平方指数核是无穷可微的,作为协方差函数,具有所有阶导数和2个自由参数:

$$k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2l^2} \|x - x'\|^2\right) \quad (10)$$

其中, l 为长度比例参数,其控制函数改变的水平尺度, σ_f^2 为输入信号的噪声方差参数,其控制函数的垂直尺度变化。

在函数假设中,长度比例参数 l 起着重要作用。

较小的 l 会使函数曲线变化更快,且协方差随着点间距离的增大而迅速衰减到零。相反,较大的 l 会导致缓慢的变化,会产生一个非常平滑的函数。在实际训练过程中,为了模拟快速变化的函数,应该使用较小的 l ;要外推一个远离训练数据的值,应选择较大的 l 。噪声方差参数 σ_f^2 决定了函数与其平均值之间的平均距离,该距离通常用作衡量先验函数簇波动的尺度因子。

2.2.2 有理二次核(RQ)

有理二次核可以看作是将不同长度比例的平方指数核相加,在一定程度上可以替代平方指数内核,解决高斯核计算速度慢的问题。它有2个自由参数,函数形式为

$$k_{RQ}(x, x') = \left(1 + \frac{\|x - x'\|^2}{2\alpha l^2}\right)^{-\alpha} \quad (11)$$

其中, α 是形状参数,其定义了局部变化的速度, α 越大局部变化则越快。 l 与平方指数核的含义相同。有理二次核更适合对非光滑、粗糙的函数进行建模。

2.2.3 周期性内核(PER)

周期性内核允许模拟周期性函数,类似于平方指数核,但其距离是实际距离的正弦。周期性内核有2个自由参数,函数形式为

$$k_{PER}(x, x') = \exp\left(-\frac{\sin^2(\pi(x - x')/p)}{2l^2}\right) \quad (12)$$

其中, p 是周期参数,控制着函数上的重复结构,周期参数 p 越大,变化趋势越平缓。同时,它会导致精确的重复结构,但在实际数据中出现的重复结构通常不会有精确的振荡。 l 与平方指数核的含义相同。

SE、RQ和PER共有的一个重要特性是平稳性。核函数具有平移不变性,其中函数的值仅取决于数据点之间的差异,而不取决于数据点本身。另一个重要方面是内核的局部性,SE和RQ核函数在相似数据点之间表现出很强的协方差,但协方差随着差异的增加而衰减为零。这是一个局部内核的指示器,只显示与附近点的高度相似性。与SE和RQ核不同,PER不具有局部特征。当进行外推时,会退回到没有数据点的区域,所以核的局部特征非常重要。否则,所有外推数据点与其他数据点之间的相似度将为零,GP模型将仅由均值函数控制^[12]。

通常在模型的训练中寻找最优参数,本文采用最大化边际对数似然法。通过组合核函数中的每个超参数来建立超参数矢量 θ 。设计了训练样本的条件概率信息似然函数,并将边缘对数似然表示为式(13)。同时,采用极大似然方法最大化在这些超参数下 y 出现的概率,并根据训练样本求解超参数矢量 θ 。一旦给

出了核函数、核函数的参数和训练数据,就可以唯一确定模型。

$$L(\theta) = \log p(y|x, \theta) = -\frac{1}{2}y^T K_{yy}^{-1}y - \frac{1}{2}\log |K_{yy}| - \frac{N}{2}\log(2\pi) \quad (13)$$

2.3 基于高斯过程的高空风场预测模型建立

给定一些训练数据,高斯过程可以通过获得其后验分布来进行回归。本文建立模型的主要步骤如下:

- (1) 输入实际观测到的风速时间序列;
- (2) 确定实验的样本数量并进行单一核函数实验;
- (3) 选择合适的组合核函数;
- (4) 采用极大似然法并结合数据分布获得适合该时间序列的超参数。

2.3.1 实验数据说明

目前,ADS-B设备的覆盖仍然不完整,航班数量较少,在某些时段和高度缺乏数据。通过对各位置风场时间序列的分析,发现数据缺失程度严重。如果按照通常的风场补盲处理,可能会得到不理想的值,甚至会改变真实风场的特征和变化规律,从而影响模型的训练结果和预测的准确性。为保证所研究模型适合真实的反演风场且不受插值误差的影响,本文选取数据完整的风场序列进行短期的风场预测模型研究。

2.3.2 核函数的选择与超参数的确定

常用的核函数只适用于简单的场合,核函数的一个优点是可以组合形成一个更专门的核函数。这允许在先验模型中加入更多信息,以使预测更准确。本文的高空风场信号中,特别是在风场的拟合和外推方面,性能不足。因此,有必要深入分析每个核函数的特性,通过核函数之间的简单运算,如加法、乘法和指数运算,创建更复杂的核函数。本文考虑了多个核函数的线性组合,核函数的组合形式为

$$K_m = \sum_m a_m k_m \quad (14)$$

其中: K_m 为第 m 个简单核函数, a_m 为权重系数。GPR的核函数应该充分展示输入之间的相似性。

本文对风场的水平分量和垂直分量进行多次实验,最后为这两个分量选择了不同的组合核函数。对于水平风分量,选取有理二次核和周期性内核的组合作为核函数,组合形式如式(15)所示。对于垂直风分量,选取平方指数核与周期性内核的组合作为核函数,组合形式如式(16)所示。

$$K_X(x, x') = k_{RQ}(x, x') + k_{PER}(x, x')$$

$$= \left(1 + \frac{\|x - x'\|^2}{2\alpha l_{RQ}^2}\right)^{-\alpha} + \exp\left(-\frac{\sin^2(\pi(x-x')/p_X)}{2l_{PER}^2}\right) \quad (15)$$

$$K_Y(x, x') = k_{SE}(x, x') + k_{PER}(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2l_{SE}^2} \|x - x'\|^2\right) + \exp\left(-\frac{\sin^2(\pi(x-x')/p_Y)}{2l_{PER}^2}\right) \quad (16)$$

确定了核函数后,对模型进行训练和优化,通过最大化边缘对数似然法确定每个组合核函数的超参数,并获得更好的参数值。

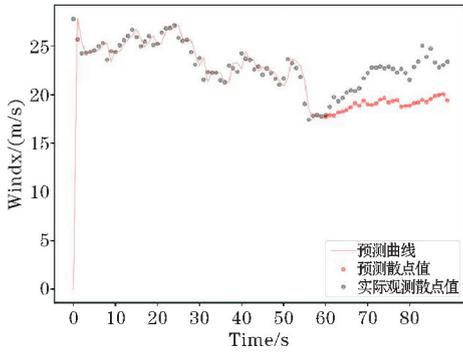
基于高斯过程的预测算法在确定核函数时,选择了两个基本核函数的加性组合,得到理想结果。未来,将会考虑核函数的自动选择算法来比较和分析更多的核函数,并尽可能详细地选择最佳的时间序列预测模型^[13]。

3 实验结果与分析

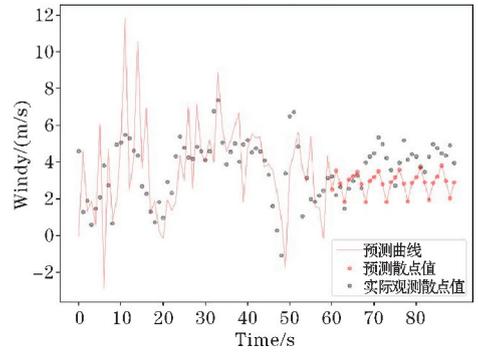
原始风场时间序列包含1 h的风场数据,通过反复实验,并结合现有数据的特点,本文利用1 h数据进行实验。控制样本点为120,将样本点分为训练集A1[1:60]、验证集A2[61:90]和测试集A3[91:120]。在实际预测中,通过前60个采样点预测接下来30个采样点的风场,即利用30 min的数据集预测未来的15 min。将选用2018年1月1-10日的完整风场时间序列来描述实验过程,并分别对SARIMA模型和高斯过程回归模型的预测结果进行对比分析。

3.1 基于SARIMA模型的预测结果

将SARIMA模型用于预测2018年1月1-10日的完整风场时间序列。为保证所研究的模型适用于实际的反演风场且不受插值误差的影响,随机选择4个位置Point1(-60, -120, 11), Point2(0, -120, 11), Point3(60, -120, 11)和Point4(60, 60, 1)的风场序列进行短期风场预测模型研究和实验。图5~8显示了选取前60个采样点预测未来30个点的值,从图中的结果来看,SARIMA模型基本上可以捕捉到短时风场的变化。特别是对于缓慢变化的 x 分量,该模型很适合于训练集,但个别值存在突变,同时,它在测试集上具有一定的泛化能力,能够基本预测风场时间序列的变化趋势。对于数据波动较大的 y 分量,结果不太理想,误差较大。首先,该模型在测试集中误差较大,个别数值的突变情况更加严重,但也保持了整体的变化趋势。预测结果偏差较大,不在可接受范围内。

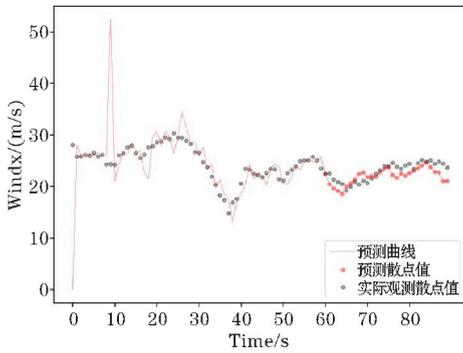


(a) Point1 x 分量

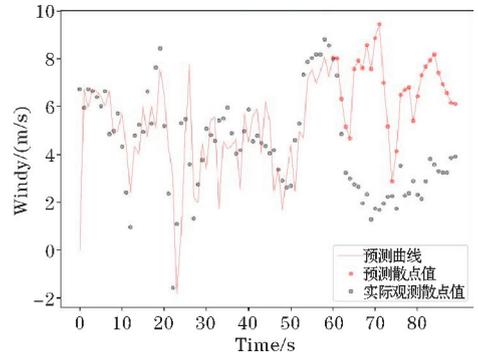


(b) y 分量

图 5 Point1(-60,-120,11)一阶差分序列 ACF 和 PACF 图

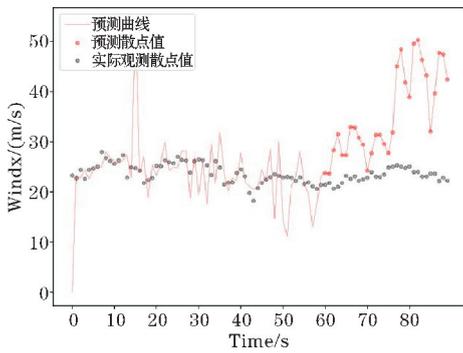


(a) x 分量

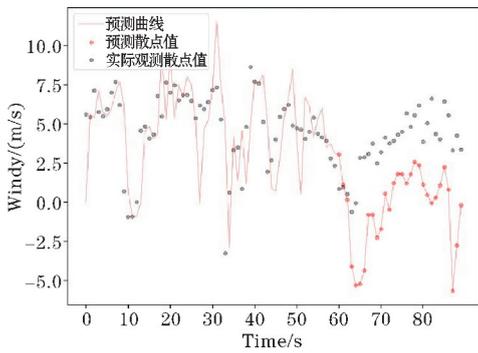


(b) y 分量

图 6 Point2(0,-120,11)一阶差分序列 ACF 和 PACF 图

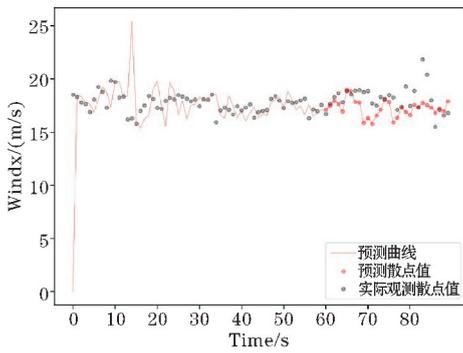


(a) x 分量

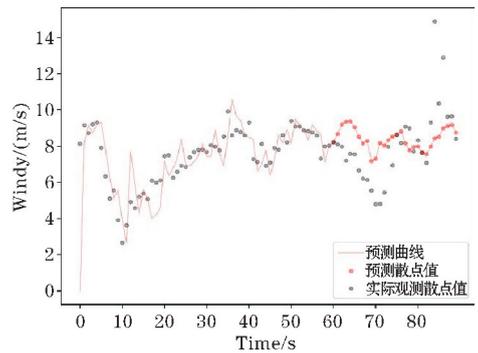


(b) y 分量

图 7 Point3(60,-120,11)一阶差分序列 ACF 和 PACF 图



(a) x 分量



(b) y 分量

图 8 Point4(60,60,1)一阶差分序列 ACF 和 PACF 图

此外,表 1 是预测的相关指标,风速预测偏差保持在 10 m/s 以内,整体结果可接受。COR 衡量的方向相似性较低,尤其是 y 分量,存在完全不相关的情况。用

R 衡量的相似性相对更可靠,基本上超过 70%。综合评价指标 Combine 的值基本在 50% 以上,低于 50% 的部分效果并不理想。

表1 基于 SARIMA 模型的预测结果准确性

位置	目标分量	MAE/ (m/s)	RMSE/ (m/s)	COR	R	Combine
454	Windx	2.93	3.19	0.83	0.97	0.90
466		1.31	1.56	0.67	0.94	0.81
478		4.13	5.27	0.27	0.85	0.56
864		1.14	1.52	0.27	0.86	0.57
454	Windy	1.22	1.47	0.03	0.82	0.43
466		3.77	4.22	0.13	0.76	0.45
478		4.15	4.58	0.48	0.91	0.70
864		1.47	1.99	0.35	0.84	0.60

3.2 基于高斯过程模型的预测结果

将高斯过程回归模型用于预测 2018 年 1 月 1-10 日的完整风场时间序列。上节中 4 个位置的风场序列,用于短时风场预测模型研究和实验。图 9~12 所示是选取前 60 个采样点预测未来 30 个点的值,其中浅蓝色区域是 95% 置信区间。从结果来看,无论是 x 分量还是 y 分量,该模型都能很好地拟合测试集数据,并预测未来风场序列的变化趋势。

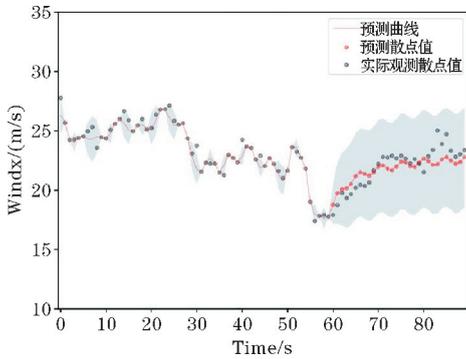
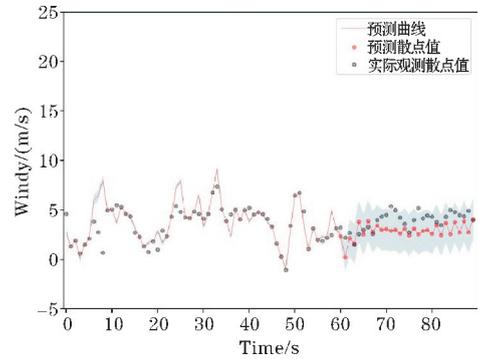
(a) x 分量(b) y 分量

图9 Point1(-60,-120,11)一阶差分序列 ACF 和 PACF 图

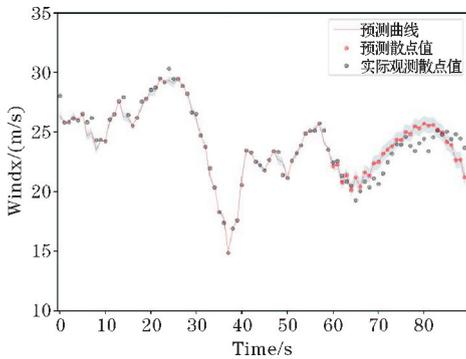
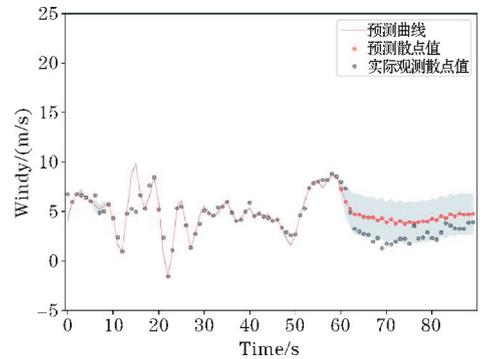
(a) x 分量(b) y 分量

图10 Point2(0,-120,11)一阶差分序列 ACF 和 PACF 图

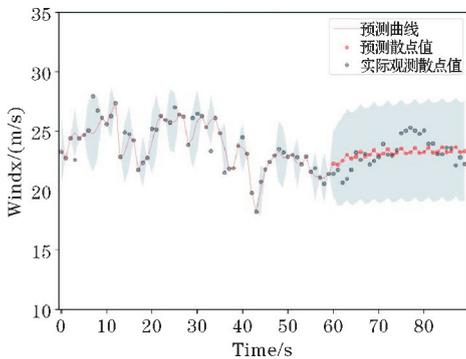
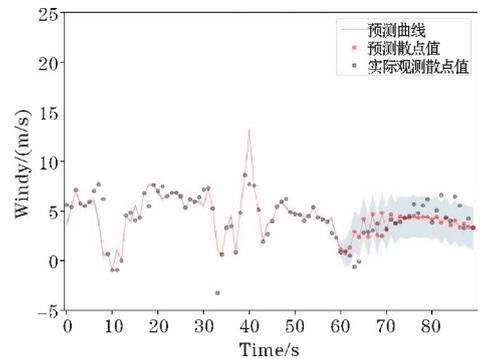
(a) x 分量(b) y 分量

图11 Point3(60,-120,11)一阶差分序列 ACF 和 PACF 图

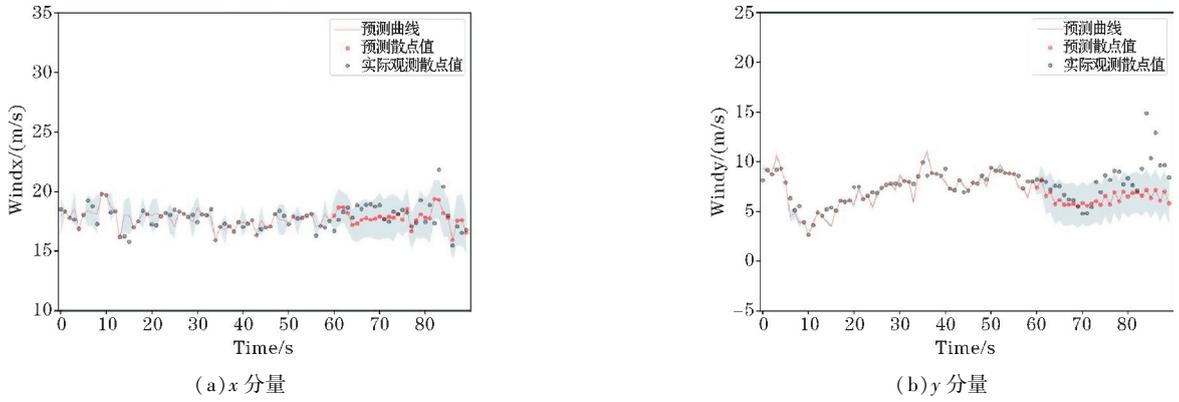


图 12 Point4(60,60,1)一阶差分序列 ACF 和 PACF 图

计算的高斯过程回归模型预测的风场序列的各项指标如表 2 所示。分别显示了 x 和 y 分量预测指标的统计结果。平均绝对误差 MAE 在 5 m/s 以内, x 分量

的准确性普遍优于 y 分量。COR 和 R 指标几乎都在 50% 以上, 综合评价指标 Combine 的值都高于 50%, 说明预测结果具有较高的可靠性。

表 2 基于 GPR 模型的预测结果准确性

位置	核函数	目标分量	MAE/(m/s)	RMSE/(m/s)	COR	R	Combine
454	Kernel 5	Windx	0.82	0.98	0.89	0.98	0.94
466			1.16	1.36	0.78	0.95	0.87
478			0.88	1.08	0.52	0.92	0.72
864			0.83	1.02	0.55	0.93	0.74
454	Kernel 1	Windy	0.85	1.03	0.25	0.86	0.55
466			2.25	2.42	0.75	0.85	0.80
478			1.13	1.50	0.65	0.95	0.80
864			1.40	1.96	0.55	0.86	0.70

3.3 预测结果对比分析

通过对两种预测方法的研究,发现基于高斯过程的回归预测模型的性能优于基于 SARIMA 的预测模型。SARIMA 模型对 x 分量有很好的预测效果,但对 y 分量的预测效果不理想。无论是 x 分量还是 y 分量, GPR 模型的预测准确性都优于 SARIMA 模型。

首先, GPR 模型能够更好地掌握一些细节信息和数据变化趋势,具有很好的泛化能力,而 SARIMA 模型更能预测序列的变化趋势,因此在一定程度上降低了准确性。其次,由于本文研究的是短时风场预测,采样点数量较少,且 GPR 模型的核函数相对简单,两种模型所花费的时间差别不大。最后,统计发现 SARIMA 模型的整体稳定性不如 GPR 模型。前者能更好地预测某些时间序列的结果,但难以预测某些趋势变化较大的序列。GPR 模型具有更好的整体稳定性和更广泛的适用性。总体而言,基于 GPR 的方法优于基于 SARIMA 的方法。

4 结束语

本文分析了高空风场时间序列的分布特征,研究

高空风场时间序列的短时预测算法。首先,建立了 SARIMA 预测模型,并根据时间序列的季节性和平稳性选择了模型细节和参数。其次,为 GPR 预测模型选择合适的核函数,并不断优化模型的超参数。最后,利用两种模型对高空风场的完整时间序列进行了预测,并对两种模型在实际时间序列中的应用结果进行了比较和评价。比较分析了上述模型的优缺点和适应性,得出了适合高空风场序列预测的模型。

研究表明,基于 GPR 的方法优于基于 SARIMA 的方法。对于具有明显规律性的长期时间序列问题,基于 GPR 的方法比基于 SARIMA 的方法更可取。此外,还应考虑结构建模的性能,因为与 SARIMA 相比, GPR 需要大量的训练时间。计算时间已成为所提出的 GPR 方法的主要缺点,但这一缺点对本文所研究的问题没有影响,时间差异不大。当然,在与大数据预测相关的其他问题上,尽管 GPR 在预测性能上优于 SARIMA 模型,但这个计算问题是使用 GPR 方法取代 SARIMA 模型的最大障碍。

参考文献:

[1] 崔世锋,程正飞,吴月友. 安徽省安庆地区高空

- 风特征分析[J]. 安徽农业科学, 2010, 38(22): 11803-11805.
- [2] 黄成涛, 王立新, 钟伯文. 高空风对大型客机航线性能的量化影响[J]. 北京航空航天大学学报, 2017, 43(7): 1348-1354.
- [3] Zhu J, Wang H, Li J, et al. Research and Optimization of Meteo-Particle Model for Wind Retrieval [J]. Atmosphere, 2021, 12(9): 1114.
- [4] 陈天舒. 基于 ARIMA 与 GPR 组合模型的人民币汇率预测[D]. 济南: 山东大学, 2015.
- [5] Swastanto B A. Gaussian process regression for long-term time series forecasting [D]. Delft: Technische Universiteit Delft, 2016.
- [6] 徐俊, 徐文辉, 曾鑫, 等. 基于大数据挖掘电量预测方法的创新及应用[J]. 电力大数据, 2018, 21(10): 57-61.
- [7] 庄景璞. 基于神经网络的 SARIMA 模型人才需求与培养匹配预测研究[D]. 南京: 南京邮电大学, 2020.
- [8] 王晓. 基于时间序列的铁路客流量预测研究 [D]. 西安: 长安大学, 2020.
- [9] 张亚婕. 基于 ARIMA 模型对股票和指数预测结果的简单比较分析[J]. 市场研究, 2019(11): 23-26.
- [10] Roberts S, Osborne M, Ebden M, et al. Gaussian processes for time-series modelling [J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2013, 371(1984): 20110550.
- [11] Rasmussen C E, Williams C K I. Model selection and adaptation of hyperparameters [M]. Gaussian processes for machine learning. MIT Press, 2005: 105-128.
- [12] Cai H, Jia X, Feng J, et al. Gaussian Process Regression for numerical wind speed prediction enhancement [J]. Renewable energy, 2020, 146: 2112-2123.
- [13] Roman I, Santana R, Mendiburu A, et al. Evolving Gaussian process kernels from elementary mathematical expressions for time series extrapolation [J]. Neurocomputing, 2021, 462: 426-439.

Research on Short-term Prediction Method of High-altitude Wind Field based on ADS-B

ZHEN Fudong¹, ZHOU Shuyue², CHEN Lijing¹, ZHU Jiahui², WANG Haijiang²

(1. Gansu Sub-bureau of Northwest Air Traffic Management of CAAC, Lanzhou 730087, China; 2. Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: The use of aircraft as weather sensors is the latest development in air traffic management and meteorological research. Aircraft data will help to improve the accuracy of meteorological information monitoring and forecasting, which will be a major breakthrough in the field of aviation meteorology in China. Combined with the wind field inversion results of ADS-B and ModeS data, the richer data set makes the prediction effect of short-term wind field better than the traditional methods. In this paper, a new, multi-functional and high-precision optimized wind field data is used to predict the short-term high-altitude wind field. In order to improve the accuracy of short-term prediction of mesoscale high-altitude wind field, this paper studies the short-term prediction algorithm of high-altitude wind field based on high-precision and high-coverage wind field inversion data. By analyzing the distribution characteristics and changing trend of time series of high-altitude wind field, two time series prediction models of SARIMA and GPR are established. The advantages and disadvantages of the two models and their application directions are discussed respectively, and the prediction accuracy of the model is quantitatively analyzed. The results show that the method based on GPR is better than the method based on SARIMA. In the acceptable forecast time range, the prediction model based on GPR can better capture the changing trend of high-altitude wind field and has better generalization ability. In the case of less available data, the complexity of the GPR model will not be too high, and the time will be relatively less, that is, less time in exchange for higher accuracy, which is very suitable for the short-term prediction of high-altitude wind fields.

Keywords: automatic dependent surveillance broadcast; secondary surveillance radar; SARIMA; GPR