

采用 k 近邻进行空间相关性超短期风速预测

杨正瓴¹, 赵 强¹, 吴炳卫¹, 侯谨毅¹, 陈 曦¹, 张 军²

(1. 天津大学 电气自动化与信息工程学院, 天津 300072;

2. 天津大学 天津市过程检测与控制重点实验室, 天津 300072)

摘要:提高超短期风速预测准确率和可靠性的途径之一,是从历史观测值中充分挖掘风速相关性的特征和规律。将本地最新的风速历史观测值结合按照最优延迟时间提前上游风速观测值,形成空间相关性 k 近邻预测的参考矢量;以相关系数作为相关性的具体评价指标,从风速历史观测值中优选出该参考矢量的 k 个最相似的近邻;采用 7 种回归模型进行本地的未来风速预测。荷兰 Huijbertgat 地区冬季风速预测的仿真结果表明:使用线性回归、偏最小二乘回归、最小二乘支持向量回归 3 个优化模型预测,优化的 k 近邻数量为 100 左右,优化的历史数据年数为 10 a;空间相关性 k 近邻风速预测能够有效使用历史数据的相似性进行可靠的超短期风速预测。

关键词:风速预测;超短期;空间相关性; k 近邻;历史观测值

中图分类号: TM 715; TM 614

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2019.03.028

0 引言

风能是可再生清洁能源。风电功率预测是能源互联网、智能电网、电力系统与电力市场安全高效运行的基础信息^[1-2]。风电功率预测的时间尺度可分为中长期、短期和超短期。中长期预测多用于机组检修计划、风电场规划和年度发电计划等;准确可靠的短期预测,对减少弃风、优化发电计划与备用等具有重要意义;超短期预测则是优化调频及旋转备用容量以及在线优化与调度的基础数据。提高风电功率预测的准确率和可靠性,是“十三五”风电发展的主要任务之一。

风电功率预测的基础信息包括风速和风向等的未来预测时间序列、风电场实际出力的历史量测时间序列、风速和风向等的历史观测资料、风力发电机的“风速-功率”转换关系(风速到风电功率的非映射关系)等^[1-3]。由于风力发电机的机械惯性等作用,并且实际中的风速风向在不断变化,这使得“风速-功率”转换关系经常呈现为一条有一定宽度的带子,而不是一条精确的单值曲线。因此,当前提高风电功率预测的最新思路是“先预测风速,再预测风电转换,可以分开处理这 2 种不同机理的不确定性”^[1]。即先可靠地预测风速等,再可靠地提高功率转换环节的准确率与自适应能力^[2-3]。

现有的风电预测技术主要有 2 类:物理方法是采用天气预报得到的未来风速风向,再通过“风速-功率”曲线得到未来的风电功率预测时间序列;统计学方法是采用风速预测时间序列和风电场实际出力

的历史量测时间序列,建立各种统计学模型以得到未来的风电功率值^[1-7]。

在数值天气预报(NWP)方面,“全球模式”预报的效果大约从 2011 年开始进入相对饱和期,如欧洲中期天气预报中心 ECMWF(European Centre for Medium-range Weather Forecasts)列出的对位势高度(对应大气压)、温度、风速等参数的预报效果均开始徘徊不前。在统计学方面,只利用本地历史数据的统计学方法对未来的预测一般在 12 h 以内。由于上述原因,从 2014 年以来,空间相关性预测受到国内外的广泛重视^[8-12]。

空间相关性预测的基本含义,是利用被预测地点(目标站点,下文简称“本地”)周边地区的风速风向等信息来预测本地未来的风速风向等。这些周边地区的风速风向等信息,既可以是当前和历史的观测值,又可以是对未来的预测值。空间相关性预测既可以采用各类回归等统计学方法,也可以引入流体力学方程等进行物理方法的预测^[10],其可用于中长期、短期和超短期风速风向预测^[8-11]。文献[1,9]对空间相关性预测研究进行了综述,故而本文不再对其进行一般性的回顾。目前基于空间相关性的风电功率预测尚未形成一套完整的理论体系^[9],在适用条件、相关性分析指标、回归模型、历史资料使用等方面都有提高的空间。

我国的风电功率预测还有自身的特殊性^[1-2,12]:一是我国大多数地区天气预报的难度往往超过欧美地区;二是我国的“风速-功率”曲线明显比欧美复杂。进一步提高我国的风电功率预测效果,应该立足于解决这 2 个特殊性问题。对于第一个问题,在我国季风区,充分利用冬季风、夏季风的空间相关性是一个值得重视的研究方向^[1,12]。作为我国“十

收稿日期:2018-05-19;修回日期:2018-12-23

基金项目:国家自然科学基金资助项目(U1766210)

Project supported by the National Natural Science Foundation of China(U1766210)

三五”风电发展重心的东南部以及沿海地区,恰好是季风的主要路径,其风速相关性的空间距离可以超出 2 000 km,延迟时间超出 27 h,明显超过了欧美常见的 50 km、4 h 的典型范围。

本文不对风速的空间相关性预测进行全面改进,而将重点集中解决一些不足之处,如“未充分利用本地数据序列在不同时间窗口之间的相关性”、“仅以当前时刻之前有限一段窗口的实测数据为基础难以全面反映空间相关性”^[1]等,即使用历史资料来进一步提高空间相关性预测的可靠性和准确性。本文将统计学方法中的 k 近邻算法^[13-15]引入空间相关性的超短期风速预测中,并对预测模型的阶数、训练模型的历史风速样本容量与预测误差的关系进行数值实验/仿真研究。

1 空间相关性风速预测的要点

1.1 共有要点的扼要回顾

空间相关性预测是一种独立于 NWP 的方法,具有弥补 NWP“家族性缺陷”^[2]的特殊潜力。

目前利用空间相关性预测风速的常见思路如文献[1]中图 1 所示,其包含 2 个重要的步骤:第一步,根据风向和风速的相关性,从周边地区优选出可以作为本地预测上游的地点;第二步,根据这些上游地点的风速等信息,预报出本地的未来风速。预报模型除了常见的统计学模型之外^[16],还可以采用物理方法(如采用 Navier-Stokes 方程组^[10])。

1.2 空间相关性的特殊优点

a. 具有提高大风预报效果的特有潜力。

显然,当参考站与目标站之间天气相关性不明显,处于低风速或者模式匹配失败时,空间相关性预测会失效;相反,当本地及周边地区出现大风时,空间相关性往往能够有效地保证风速预报的效果。换言之,在大风时期,空间相关性预测是“组合预测”的可靠单项模型之一;无风、小风时应予舍去。

从预测原理和实践上看,NWP 的误差与风速大小自身没有必然的联系,亦即 NWP 不能保证大风预报得更准确^[17]。

b. 我国季风区特别适合采用空间相关性预测。

如文献[1,12]所述,我国典型冬季风、夏季风每年共计有 7 个月以上的时间。风速相关性范围超出 2 000 km、27 h,已经进入短期风电功率预测的时间尺度。

特别是台湾海峡区域,不仅风能资源丰富,而且由于福建、浙江沿海地区的地形和台湾岛形成一个长度超过 400 km 接近平行的海面,使得该区域具有很强的风速空间相关性,延迟时间达到 6 h。

c. 一定的强对流天气预报预警能力。

强对流天气的空间尺度小,水平范围一般在十

几至二三百千米,带有明显的突发性且持续时间短,一般小于一至十几小时。但往往伴随破坏力强的大风,可能引起风电功率爬坡或风机损坏。

NWP 通常难以有效预报强对流天气,我国目前实际中以气象雷达的动态监视为主。在强对流多发地区,建立较为密集的测风塔,则可用空间相关性对其进行一定程度的预测。

2 时间序列 k 近邻预测的要点

一般认为 k 近邻预测由 Cover 和 Hart 在 1967 年提出^[14]。在一维实数时间序列 $x_i (i=1,2,\dots)$ 预测时,设当前时刻为 t ,采用 k 近邻的基本步骤如下。

首先,以合适的维数 p 构造包含当前和最近点的参考矢量 $\mathbf{V}_R = \mathbf{V}_t = (x_{t-p+1}, x_{t-p+2}, \dots, x_{t-1}, x_t)$,以某种相关性或距离指标,从历史值中找到 k 个与 \mathbf{V}_R 最相似的历史矢量 $\mathbf{V}_j (j=1,2,\dots,k)$ 。

然后,从 k 个 \mathbf{V}_j 中回归/训练出式(1)所示模式的具体模型 F (如形式和参数)。这里的具体模型,既可以是有显式的解析表达式的线性回归(LR)、偏最小二乘回归(PLSR),也可以是支持向量机、人工神经网络等^[16]。

$$F(\mathbf{V}_R) = x_{t+1} \quad (1)$$

最后,将当前观测值 \mathbf{V}_t 代入该模型,即可得到 \hat{x}_{t+1} 的预报值:

$$\hat{x}_{t+1} = F(\mathbf{V}_t) = F(x_{t-p+1}, x_{t-p+2}, \dots, x_{t-1}, x_t) \quad (2)$$

对于混沌的时间序列,当采用优化的模型阶数 p 以及足够长的历史数据时, k 近邻预测的误差一般可以达到最小值。

对于常见的复杂时间序列, k 近邻预测被证实为具有预测可靠性高、准确率高,适用的数据类型范围宽,以及可以自动回避历史数据丢失或“坏”数据(如 outlier)的优良性能^[13-15]。简言之, k 近邻预测是一种能够充分利用历史数据的相似性的优秀方法,其主要缺点是计算速度偏慢,所使用的历史数据越长,计算速度会越慢,尽管预测的准确率会越高。在风电功率预测领域,近年全球能源预测竞赛中, k 近邻是优胜模型之一^[13]。

3 采用 k 近邻的空间相关性风速预测

3.1 空间相关性风速预测 k 近邻算法的原理

采用 k 近邻的空间相关性风速预测原理示意图如图 1 所示,结合本地和上游风速之间的优化延迟时间,可以将 k 近邻预测的一维时间序列模型推广为多维时间序列的模型,即风速空间相关性的 k 近邻预测。

具体地,在空间相关性风速预测 k 近邻算法中,当前的参考矢量和历史观测值中的 k 近邻不仅包含本地的风速,而且还包含经优化延迟时间 τ 调整后

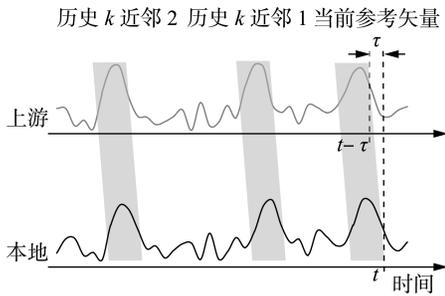


图 1 采用 k 近邻的空间相关性风速预测原理示意图

Fig.1 Schematic diagram of wind speed prediction based on spatial correlation by k -nearest neighbor

的上游各风速。如图 1 所示,在参考矢量和历史 k 近邻中,若本地时间为 t ,则上游时间为 $t-\tau$ 。

优化延迟时间 τ 可以从历史观测资料中分析计算得到,具体方法有交叉相关函数(SCCF)^[1]、小波相干等。各地风速在参考矢量中的历史矢量长度 p ,可以采用非线性时间序列中的 Takens 嵌入定理确定,即 $p = 2D_2 + 1$,其中 D_2 为风速时间序列的维数^[18]。

由于风在传播过程中是时变的,所以先通过历史 k 近邻建立回归预测模型,再将上游和本地风速最近的观测值以当前参考矢量的方式代入预测模型,以预报经过延迟时间 τ 后本地的未来风速。

3.2 空间相关性风速预测 k 近邻算法的流程

图 2 为预测时间段中荷兰 Huibertgat 及其上游 Lauwersoog、Stavoren、Ijmuiden 和 Cadzand 地区的风速。该段风速序列取自 2013 年 12 月 2 日至 7 日,为当月大风的时期。以荷兰的 Huibertgat 为被预测地点(本地),在冬季时期,通过历史观测分析可得,其西南方向的 Lauwersoog 为最接近提前 1 h 的上游地点,Stavoren 提前约 2 h, Ijmuiden 提前约 3 h, Cadzand 提前约 4 h,并且都具有较高的相似性。

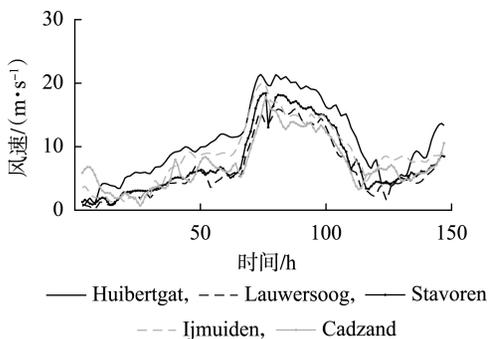


图 2 荷兰 Huibertgat 及其上游 Lauwersoog、Stavoren、Ijmuiden 和 Cadzand 的风速

Fig.2 Wind speeds of Huibertgat and its upstream Lauwersoog, Stavoren, Ijmuiden and Cadzand in Holland

记 Huibertgat 的风速为 v_{Ht} , Lauwersoog 的风速为 x_{Lt} , Stavoren 的风速为 x_{St} , Ijmuiden 的风速为 x_{It} , Cadzand 的风速为 x_{Ct} ,其中 $t = \dots, -2, -1, 0, 1, 2, \dots$ 。记 t 为当前时刻, Huibertgat 的未来风速 $\hat{v}_{H(t+1)}$ 、

$\hat{v}_{H(t+2)}$ 、 $\hat{v}_{H(t+3)}$ 、 $\hat{v}_{H(t+4)}$ 是被预测量。

图 3 为采用 k 近邻的空间相关性风速预测流程图。在对 $\hat{v}_{H(t+1)}$ 进行预测时,不妨设预测模型的阶数 $p=3$,则对应当前的参考矢量为 $\mathbf{R} = (x_{L(t-3)}, x_{L(t-2)}, x_{L(t-1)}, v_{H(t-2)}, v_{H(t-1)}, v_{H(t)})$,即图 1 最右边阴影所示的四边形。依此矢量的模式,采用 Pearson 相关系数,可以从 Huibertgat 和 Lauwersoog 的历史观测风速时间序列中找出 k 个最相似的历史矢量,如 $\mathbf{L}_i = (x_{L(i-3)}, x_{L(i-2)}, x_{L(i-1)}, v_{H(i-2)}, v_{H(i-1)}, v_{H(i)})$ 和 $\mathbf{L}_j = (x_{L(j-3)}, x_{L(j-2)}, x_{L(j-1)}, v_{H(j-2)}, v_{H(j-1)}, v_{H(j)})$ 等,这里 i, j 都表示比 t 更早的时刻,后同。



图 3 采用 k 近邻的空间相关性风速预测流程图

Fig.3 Flowchart of wind speed prediction based on spatial correlation by k -nearest neighbor

采用线性回归和偏最小二乘回归的预测方程为:

$$\hat{v}_{H(t+1)} = a_L + a_{L0}x_{L(t-3)} + a_{L1}x_{L(t-2)} + a_{L2}x_{L(t-1)} + a_{L3}v_{H(t-2)} + a_{L4}v_{H(t-1)} + a_{L5}v_{Ht} \quad (3)$$

其中, $a_L, a_{L0} - a_{L5}$ 为回归方程 $\mathbf{AL} = \mathbf{V}$ 的系数,为方便,记 $\mathbf{A} = (a_L, a_{L0}, a_{L1}, a_{L2}, a_{L3}, a_{L4}, a_{L5})$,这些系数由回归方程得出(参照式(1)), $\mathbf{L} = (\mathbf{L}_i, \mathbf{L}_j, \dots)$ 由 k 个最相似历史矢量构成, $\mathbf{V} = (v_{i+1}, v_{j+1}, \dots)$ 。

对于支持向量机、神经网络,则由 $\{\mathbf{L}, \mathbf{V}\}$ 构成训练集,以得到相应的预测模型。

对 $\hat{v}_{H(t+2)}$ 进行预测时,将 Stavoren 的风速 x_{St} 按照上述模式加入参考矢量、历史矢量中,依此类推,则可以实现对 Huibertgat 的未来风速 $\hat{v}_{H(t+1)}$ 、 $\hat{v}_{H(t+2)}$ 、 $\hat{v}_{H(t+3)}$ 、 $\hat{v}_{H(t+4)}$ 的预测。

空间相关性 k 近邻算法的思路与“离线分类优化模型,在线确认空间相关性并匹配预测模型”^[11] 有一定的类似性,也可以将上述风速观测值扩展到

使用气压、风向、温度、湿度等更多的气象条件,并可望得到更好的预报效果。

4 算例仿真及分析

“蒙古高压”是横亘欧亚大陆的半永久性冷高压,通常 9 月形成,在其后长达 7 个月左右的时间控制着欧亚大陆的气候变化。因此,我国冬季的风速变化和欧洲冬季的风速具有较为明显的相似性。如文献[19]中图 1 表明,德国北部某地冬季风速的相关性明显高于夏季。典型冬季风时期,荷兰风速和我国冬季风(以天津为例)的风速历史观测时间序列具有相近的最大 Lyapunov 指数,即 $\lambda_1 \approx 0.07$,时间窗口上的相关维数 $D_2 \approx 1$ ^[20],因此本文采用荷兰的多地风速进行示例。荷兰各地风速都是 1 h 平均风速,共计 18 a 的历史观测值。

采用逐点向前的预测,即在时刻 $t=0$ 预测未来的 4 点风速 $\hat{v}_{H1}, \hat{v}_{H2}, \hat{v}_{H3}, \hat{v}_{H4}$,并由它们计算出平均误差、均方根误差等;在时刻 $t=1$,预测未来的 4 点风速 $\hat{v}_{H2}, \hat{v}_{H3}, \hat{v}_{H4}, \hat{v}_{H5}$,并计算平均误差、均方根误差等;依此类推。并且以回归模型阶数 p 和 k 近邻的数量为参数进行循环,得到预测误差随参数 p 和 k 的变化规律。

图 2 所示预测时间段中风速变化较大,并且观测值中没有明显的坏数据,5 个风速时间序列均采用双正交小波 bior 6.8 进行了 1 层去噪。如果仅采用 Huibergat 本地的观测值,采用文献[18]的偏最小二乘回归、最小二乘支持向量回归(LSSVR)方法,提前 1~12 h 风速预测的均方根误差见表 1,其中每步间隔 1 h,后同。表中均方根误差超过了我国季风区风速的预测误差,这是由于荷兰没有典型的季风现象,其风速的空间相关性明显低于我国。

表 1 仅采用本地风速数据对荷兰 Huibergat 未来风速多点预测的均方根误差

Table 1 Root mean square error of multi-point wind speed prediction of Huibergat, Holland only by local wind speeds

提前 步数	均方根误差/($\text{m}\cdot\text{s}^{-1}$)			
	PLSR		LSSVR	
	直接式	迭代式	直接式	迭代式
1	0.30	0.30	0.30	0.30
2	0.66	0.33	0.61	0.34
3	0.93	0.93	0.87	0.92
4	1.14	1.08	1.08	1.07
5	1.34	1.54	1.28	1.49
6	1.53	1.70	1.49	1.65
7	1.72	2.04	1.68	1.96
8	1.90	2.24	1.87	2.17
9	2.08	2.40	2.06	2.33
10	2.26	2.60	2.24	2.54
11	2.43	2.72	2.43	2.67
12	2.61	2.87	2.61	2.83

在空间相关性 k 近邻算法中,采用线性回归、偏

最小二乘回归、最小二乘支持向量回归、广义回归神经网络(GRNN)、随机森林回归(RFR)、BP 神经网络和 RBF 神经网络等作为回归预测的模型。

图 4 为采用预测年份 2013 年之前 2 a 的历史风速最相似的 80 个 k 近邻,用偏最小二乘回归对荷兰 Huibergat 未来风速提前 1—4 步的预测结果。

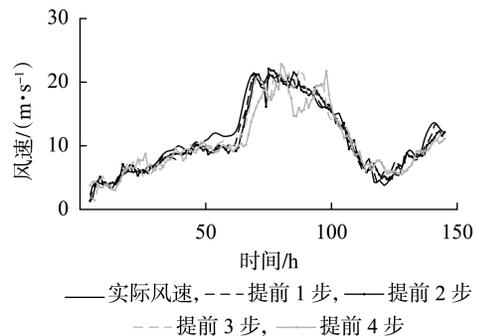


图 4 荷兰 Huibergat 风速提前 1—4 步的预测序列

Fig.4 Wind speed predicted series of Huibergat, Holland with 1 to 4 hours ahead

从图 4 可见,预测提前的时间越长,预测结果越偏离实际值,即如果当前时刻为 t ,对未来时刻 $t+1$ 的预测误差明显小于对未来时刻 $t+4$ 的预测误差,这是预测中的常见现象。

4.1 空间相关性风速预测的 k 近邻算法的回归模型优选

图 5 为采用预测年份 2013 年之前 2、4、6 a 的 5 地风速历史观测值,以上述 7 种回归模型预测的 Huibergat 未来风速均方根误差随 k 近邻的数量变化的曲线。

从图 5 可见,线性回归、偏最小二乘回归、最小二乘支持向量回归这 3 种回归模型在 k 较大时的预测误差大体相当;当 k 在 [90, 120] 范围内,预测的均方根误差大体处在最小值;当 k 再增大后,预测误差反而会出现轻微增大的趋势。类似文献[15]中图 4 所示。另外 4 种回归模型预测效果偏差,这是由于传统的神经网络类方法性能不稳定,泛化能力差,众所周知,支持向量机是为克服人工神经网络在小样本下的性能下降而提出的。

4.2 k 近邻预测的误差随历史观测值的年数增加而降低

由图 5 可知,线性回归、偏最小二乘回归、最小二乘支持向量回归这 3 种回归模型的预测误差大体相同。由于线性回归、偏最小二乘回归都有显式的解析表达式,可以为将来的参数自动优化等研究提供更好的基础性支持^[18],所以本文仅采用这 2 种模型进行预测误差与历史观测值的年数关系的仿真。

图 6 为以线性回归和偏最小二乘回归为回归模型,分别采用预测年份 2013 年之前 2、4、6、8、10、12、14、16 a 的 5 地风速观测值,对 Huibergat 未来风速预测的均方根误差随 k 近邻数量变化的曲线。

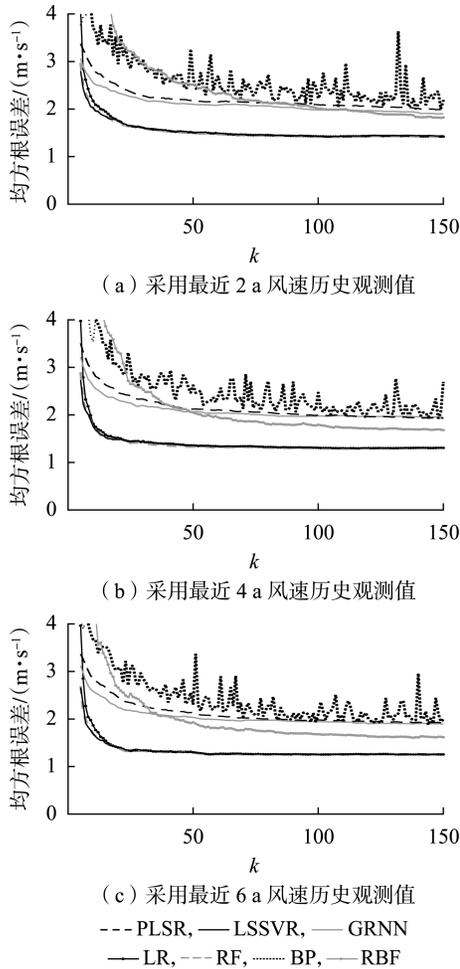


图 5 空间相关性风速 k 近邻预测的回归模型优选
Fig.5 Optimal regression model selection of wind speed prediction based on spatial correlation by k -nearest neighbor

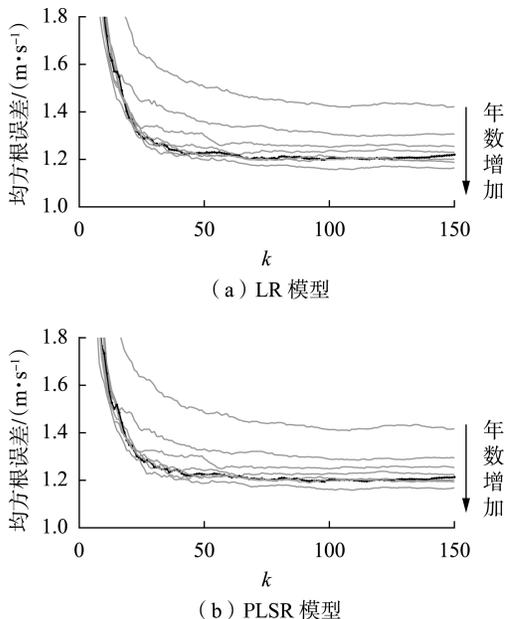


图 6 均方根误差随 k 近邻数量变化的曲线
Fig.6 Curves of root mean square error vs. k

从图 6 可见,采用预测年份 2013 年之前 10 a (如图中粗线所示)的风速历史观测值基本上可以

达到较好的预测效果。尽管进一步增加历史数据可使误差进一步减小,但减小的幅度并不明显,反而会使得计算速度进一步变慢。

图 7 为 $k=100$ 时均方根误差随年数 y 变化的函数拟合结果。

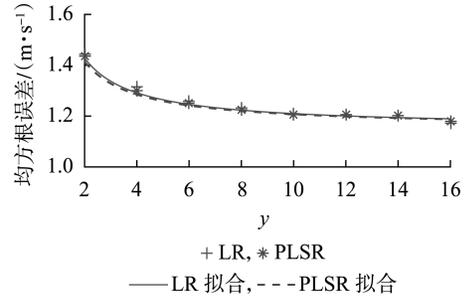


图 7 k 近邻均方根误差与年数关系

Fig.7 Relationship between root mean square error of k -nearest neighbor and number of year

经指数函数、幂函数、有理函数拟合比较后,考虑到数理统计学中样本容量作用的常见函数形式,最终的拟合函数为:采用线性回归, $e_{RMSE} = 0.546 4/y + 1.155$,拟合的均方根误差为 $0.014 44 \text{ m/s}$;采用偏最小二乘回归, $e_{RMSE} = 0.519 1/y + 1.155$,拟合的均方根误差为 $0.012 26 \text{ m/s}$ 。

采用预测年份 2013 年之前 2、4、6、8、10、12、14、16 a 的历史观测值,以线性回归、偏最小二乘回归进行预测的计算机运算时间参考值见图 8。

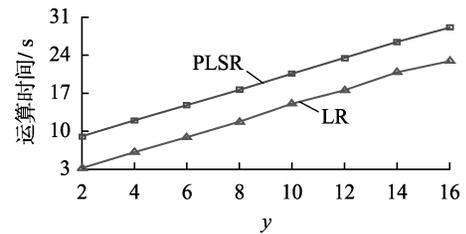


图 8 LR 和 PLSR 预测的运算时间参考值

Fig.8 Calculation time reference predicted by LR and PLSR

图 8 中运算时间随着年数的增加基本上呈现线性增大的趋势。尽管不同性能的计算机消耗不同的时间,但线性增大的趋势表明,使用长期的历史观测资料具有潜在的可能性,计算机耗时不会显著增加,但预测的可靠性会不断增大。

4.3 k 近邻预测的误差随模型阶数的变化

图 9 为采用线性回归和偏最小二乘回归模型,以预测年份 2013 年之前 2 a 的 5 地风速观测值对 Huibertgat 未来风速预测的均方根误差随模型阶数 p 的变化,这里 $p \in [2, 11]$ 。

根据 Takens 嵌入定理,预测模型的优化阶数(即相空间重构的优化阶数) $p = 2D_2 + 1$,即 p 为 $2 \sim 3$ ^[18]。考虑到天气变化 D_2 可达 $2 \sim 4$,这里示例的 p 的取值范围一般在 $2 \sim 11$ 范围内,即包含 $2 \times 4 + 1 = 9$ 。

从图 9 可见,选取模型阶数 p 为 $2 \sim 3$ 即可,再增

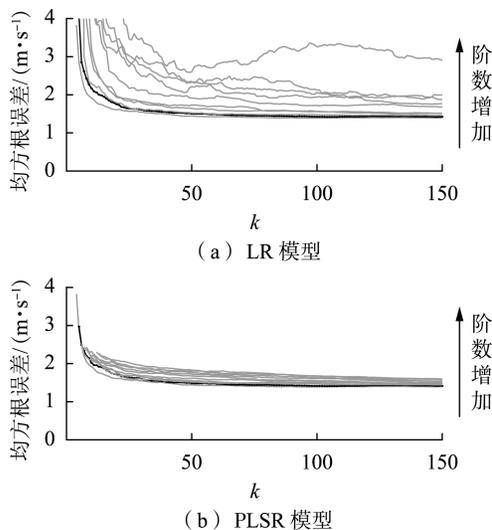


图 9 模型阶数与预测误差之间的关系

Fig.9 Relationship between model order and prediction error
大 p 会引起预测误差的增大,且降低计算的速度。

5 结论

为将本地及其上游地区风速等历史观测值时间序列中的风速相似性引入空间相关性预测,本文将本地和其上游地点的最新时间窗口上的风速合并成一个参考矢量,根据该参考矢量在本地和上游历史资料中优选 k 近邻,并采用线性回归、偏最小二乘回归、最小二乘支持向量回归、广义回归神经网络、随机森林回归、BP 神经网络和 RBF 神经网络 7 种回归模型进行风速预测。以荷兰 Huijbertgat 冬季时期的超短期风速预测为例,进行了 k 近邻数量、采用的历史数据年数的优选。主要结论如下。

可以采用相关系数作为选取 k 近邻的计算指标,采用交叉相关函数、小波相干从历史观测值中计算上游对本地优化延迟时间,采用 Takens 嵌入定理确定历史矢量的长度。

采用线性回归、偏最小二乘回归、最小二乘支持向量回归作为回归预测模型,近 10 a 的历史资料以及偏低的模型阶数较好, k 近邻数量最优为 100 左右。适度增大历史资料的年数并选择较低的模型阶数,是提高空间相关性风速预测的 k 近邻算法效果的主要手段。由于 k 近邻的相似性,一般没有必要在寻找更好的回归模型方面做更多的研究。

沿我国冬季风、夏季风主要路径建立更多的专业测风塔^[12],积累更多更可靠的历史资料,并将计算流体力学方法、大气运动方程引入 k 近邻预测,是值得进一步研究的方向。

参考文献:

[1] 薛禹胜,陈宁,王树民,等. 关于利用空间相关性预测风速的评述[J]. 电力系统自动化,2017,41(10):161-169.
XUE Yusheng, CHEN Ning, WANG Shumin, et al. Review on wind

speed prediction based on spatial correlation[J]. Automation of Electric Power Systems, 2017, 41(10): 161-169.

- [2] 乔颖,鲁宗相,闵勇. 提高风电功率预测精度的方法[J]. 电网技术,2017,41(10):3261-3268.
QIAO Ying, LU Zongxiang, MIN Yong. Research & application of raising wind power prediction accuracy[J]. Power System Technology, 2017, 41(10): 3261-3268.
- [3] 杨茂,杨琼琼. 风电机组风速-功率特性曲线建模研究综述[J]. 电力自动化设备,2018,38(2):34-43.
YANG Mao, YANG Qiongqiong. Review of modeling of wind speed-power characteristic curve for wind turbine[J]. Electric Power Automation Equipment, 2018, 38(2): 34-43.
- [4] 王兰,李华强,吴星,等. 基于改进局域 Volterra 自适应滤波器的风电功率混沌时间序列预测模型[J]. 电力自动化设备,2016,36(8):40-44.
WANG Lan, LI Huaqiang, WU Xing, et al. Wind power chaotic time series prediction model based on improved local Volterra adaptive filter[J]. Electric Power Automation Equipment, 2016, 36(8): 40-44.
- [5] 范磊,卫志农,李慧杰,等. 基于变分模态分解和蝙蝠算法-相关向量机的短期风速区间预测[J]. 电力自动化设备,2017,37(1):93-100.
FAN Lei, WEI Zhinong, LI Huijie, et al. Short-term wind speed interval prediction based on VMD and BA-RVM algorithm[J]. Electric Power Automation Equipment, 2017, 37(1): 93-100.
- [6] 欧阳庭辉,查晓明,秦亮,等. 含核函数切换的风电功率短期预测新方法[J]. 电力自动化设备,2016,36(9):80-86.
OUYANG Tinghui, ZHA Xiaoming, QIN Liang, et al. Short-term wind power prediction based on kernel function switching[J]. Electric Power Automation Equipment, 2016, 36(9): 80-86.
- [7] 李军,常燕芝. 基于 KPCA-KMPMR 的短期风电功率概率预测[J]. 电力自动化设备,2017,37(2):22-28.
LI Jun, CHANG Yanzhi. Short-term probabilistic forecasting based on KPCA-KMPMR for wind power[J]. Electric Power Automation Equipment, 2017, 37(2): 22-28.
- [8] JUNG J, BROADWATER R P. Current status and future advances for wind speed and power forecasting[J]. Renewable and Sustainable Energy Reviews, 2014(31):762-777.
- [9] 叶林,赵永宁. 基于空间相关性的风电功率预测研究综述[J]. 电力系统自动化,2014,38(14):126-135.
YE Lin, ZHAO Yongning. A review on wind power prediction based on spatial correlation approach[J]. Automation of Electric Power Systems, 2014, 38(14): 126-135.
- [10] YE L, ZHAO Y, ZENG C, et al. Short-term wind power prediction based on spatial model[J]. Renewable Energy, 2017(101):1067-1074.
- [11] 陈宁,薛禹胜,丁杰,等. 利用空间相关性的超短期风速预测[J]. 电力系统自动化,2017,41(12):124-130.
CHEN Ning, XUE Yusheng, DING Jie, et al. Ultra-short term wind speed prediction using spatial correlation[J]. Automation of Electric Power Systems, 2017, 41(12): 124-130.
- [12] 杨正瓴,杨钊,张玺,等. 基于季风提高空间相关性预测的优化延迟时间[J]. 电力系统保护与控制,2016,44(15):33-38.
YANG Zhengling, YANG Zhao, ZHANG Xi, et al. Improving optimal lag time of spatial correlation prediction by characteristics of monsoon[J]. Power System Protection and Control, 2016, 44(15): 33-38.
- [13] ZHANG Y, WANG J. K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting[J]. International Journal of Forecasting, 2016, 32(3):1074-1080.
- [14] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1):21-27.

- [15] SKRIMPAS G A, KLEANI K, MIJATOVIC N, et al. Detection of icing on wind turbine blades by means of vibration and power curve analysis[J]. *Wind Energy*, 2016, 19(10): 1819-1832.
- [16] EFRON B, HASTIE T. *Computer age statistical inference* [M]. Cambridge, UK: Cambridge University Press, 2016.
- [17] ZHAO J, GUO Z, SU Z, et al. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed[J]. *Applied Energy*, 2016, 162: 808-826.
- [18] 杨正瓴, 刘阳, 张泽, 等. 采用最近历史观测值和 PLSR 进行空间相关性超短期风速预测[J]. *电网技术*, 2017, 41(6): 1815-1822.
YANG Zhengling, LIU Yang, ZHANG Ze, et al. Ultra-short-term wind speed prediction with spatial correlation using recent historical observations and PLSR[J]. *Power System Technology*, 2017, 41(6): 1815-1822.
- [19] HAGSPIEL S, PAPAEMANNOUIL A, SCHMID M, et al. Copula-based modeling of stochastic wind power in Europe and implications

for the Swiss Power Grid[J]. *Applied Energy*, 2012, 96: 33-44.

- [20] LI X, HU F, LIU G. Characteristics of chaotic attractors in atmospheric boundary-layer turbulence[J]. *Boundary-Layer Meteorology*, 2001, 99(2): 335-345.

作者简介:



杨正瓴

杨正瓴(1964—),男,河北灵寿人,副教授,博士,主要研究方向为风电功率预测、电力负荷预测、混沌理论及其在电力系统中的应用(E-mail: zlyang@tju.edu.cn);

赵强(1992—),男,辽宁昌图人,硕士研究生,主要研究方向为风电功率预测(E-mail: zhaoqiangtju@163.com);

吴炳卫(1994—),男,安徽滁州人,硕士研究生,主要研究方向为风电功率预测(E-mail: 1264461864@qq.com)。

Ultra-short term wind speed prediction based on spatial correlation by k -nearest neighbor

YANG Zhengling¹, ZHAO Qiang¹, WU Bingwei¹, HOU Jinyi¹, CHEN Xi¹, ZHANG Jun²

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;

2. Key Laboratory of Process Measurement and Control, Tianjin University, Tianjin 300072, China)

Abstract: One approach to improve the accuracy and reliability of ultra-short term wind speed prediction is to thoroughly exploit the characteristics and laws of wind speed correlation from historical observations. The reference vector of k -nearest neighbor prediction based on spatial correlation is formed by the combination of latest local wind speed historical observations and upstream wind speed observations adjusted by its optimal lag time. The correlation coefficient is taken as the concrete evaluation index, and k most similar neighbours of the reference vector are optimally selected from the wind speed historical observations. Seven regression models are adopted for the future local wind speed prediction. The simulative results of wind speed prediction of Huibertgat, Holland in winter show that the optimal number of k -nearest neighbours is about 100 and the optimal year number of historical data is 10 a by the prediction of three optimal models, i.e. linear regression, partial least squares regression and least squares support vector machine regression, and the proposed method can effectively use the similarity of historical data for reliable ultrashort term wind speed prediction.

Key words: wind speed prediction; ultra-short term; spatial correlation; k -nearest neighbor; historical observations

(上接第 174 页 continued from page 174)

Parallel optimal allocation of distributed PV in distribution network considering active management

SUN Min¹, YU Yuan², ZENG Wei¹, PENG Chunhua²

(1. State Grid Jiangxi Electric Power Research Institute, Nanchang 330096, China;

2. School of Electrical & Automation Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: The rapid development of AM(Active Management) technology provides possibility for large-scale distributed PV(PhotoVoltaic) power generation accessing to distribution network. An optimal allocation model with the combination of AM and optimal distributed PV allocation, which takes the installation position, capacity, and AM measures as its decision variables and the maximum distributed PV energy permeability and the minimum voltage deviation as its objectives. The multi-scenario analysis based on dichotomous K -means clustering is used to deal with the uncertainties and temporal characteristics of PV output and load, which overcomes the shortage that K -means clustering scenario reduction is sensitive to the selection of initial centroid. The multi-scenario analysis and multi-objective differential evolution algorithm based on parallel computing are proposed to solve the optimization model, and the optimal PV allocation scheme and AM strategy of distribution network are obtained. The simulative results of IEEE 33-bus distribution network show that the proposed method can effectively improve the distributed PV penetration amount and ensure the power supply quality.

Key words: active management; optimal photovoltaic allocation; energy permeability; voltage deviation; parallel computing; dichotomous K -means