Journal of Southwest University (Natural Science Edition)

DOI: 10.13718/j. cnki. xdzk. 2022.05.010

# 基于 Gappy POD 和 WRF 模式的风速数据重构

小巴桑次仁<sup>1</sup>, 索朗塔杰<sup>2</sup>, 杜媛芳<sup>1,3</sup>, 拉巴次仁<sup>1</sup>

1. 西藏大学 理学院, 拉萨 850012; 2. 西藏自治区气象局, 拉萨 850011;

3. 武汉大学 经济与管理学院, 武汉 430072

摘要:由于地形较为复杂且观测站稀疏等原因导致西藏自治区那曲市中东部一带的风速重构较为困难.计算了 WRF模式在研究区域内格点处的风场数据并以此为基础计算风场的主要特征,采用优化的经典 POD 算法将该区 域仅有的 8 个稳定台站的风观测数据应用于提升模式输出数据的精度.交叉检验结果表明该方法能够提高地形复 杂的观测站稀疏区域内任意点处的模式风速重构的精度.

关键 词: Gappy POD; WRF 模式; 风速; 重构; 复杂地形

中图分类号: O175.8 文献标志码: A

文章编号: 1673-9868(2022)05-0082-07



开放科学(资源服务)标识码(OSID): 6

# Gappy POD and WRF-Model Based Wind Speed Data Reconstruction

BASANG Tseringxiao<sup>1</sup>, SUOLANG Tajie<sup>2</sup>, DU Yuanfang<sup>1,3</sup>, LHAPA Tsering<sup>1</sup>

1. School of Science, Tibet University, Lhasa 850012, China;

2. Meteorological Bureau of TAR, Lhasa 850011, China;

3. School of Economics and Management, Wuhan University, Wuhan 430072, China

Abstract: Due to the complicated topography and sparse distribution of observation stations, it is difficult to reconstruct the wind speed in the central and eastern part of Naqu, TAR. In this research, the WRF model was utilized to acquire wind data for each grid point and the data was then used to construct the basic features of the wind field in the studied region. The improved POD technique described in this paper plays a bridging role between the WRF model and the data collected by the observation stations. The cross-validation result reveals that the strategy used in this research increased the accuracy of wind speed model reconstruction at any point theareas with complicated topography and sparse observation station distribution.

Key words: Gappy POD; WRF model; wind speed; reconstruction; complex terrain

基金项目: 霍英东青年教师基金基础性研究课题(151102).

作者简介:小巴桑次仁,博士,教授,主要从事偏微分方程反问题理论的研究.

青藏高原独特、多变的气象特征是气象学的热点研究领域.针对青藏高原及其周边复杂地形下的大气 结构和地一气物理交换过程等方面的研究已获得了丰硕的成果,然而广袤的高原却由于地面气象观测站分 布较少且极端天气、自然灾害等客观原因,很多观测站数据存在缺、漏等现象,无法提供丰富的实时观测 有效数据.因此如何提高观测数据质量,用少量的不连续的观测数据修补缺失数据是基础数据研究领域非 常重要的科学问题.本文基于严重不足的气象数据,采用 POD(proper orthogonal decomposition,本征正 交分解)优化后形成的 Gappy POD 算法与模式相结合的方式重构了近地面水平风速.

本文研究区域(图1)为中国西藏那曲地区中东部一带,平均海拔超过4500m,具有典型的复杂高山地 形特色.该区域常年大风天气,且由于地形原因风场呈现复杂结构以及高度空间变异性特点<sup>[1]</sup>.这些因素 使得中尺度模型<sup>[2]</sup>的计算变得极具挑战性.除去地形复杂的特点之外,该广袤区域内气象观测站较少且分 布极度稀疏,导致数据分析的经典方法和基于大气运动方程的数值算法无法对气象要素进行高精度预测或 重构.本文采用模式与观测数据相结合的思路,利用美国国家大气研究中心(national center for atmospheric research, NCAR)的中尺度 WRF(weather research and forecast)模式<sup>[3-5]</sup>计算了研究区域内格点数 据并以此为基础进一步计算了区域内宏观风场特征以及风速特征;本文建立的 Gappy POD 算法的主要优 势在于实现了利用局部观测数据调整宏观气象特征所存在的系统误差从而重构所得气象要素,兼具宏观大 气运动规律和局部观测数据精度高的特点.



# 1 数据、模式和算法

若研究区域内缺失数据较为严重则一般空间插值或外推方法无法较好地处理缺失值问题或重构问题. 比如 POD 是一种用于提取离散数据特征信息的降维技术,而缺失值较多时该方法在一定程度上失效.本 文通过优化原经典 L<sup>2</sup> 范数达到弥补 POD 方法在缺失数据下的不足.为了提高气象要素场的重构精度,本 文在研究区域内利用 WRF 模式<sup>[6]</sup>计算了相应气象要素在所有格点处的输出值从而建立了风速场的 POD 基,即代表该区域的宏观气象特征.观测数据在本文中视为真值(近似为零误差).利用观测数据矫正 POD 基的精准度,从而反演区域内任意一点处的气象要素值.文中所采用的思想属于如今很多科学家提倡的 "弱机理"思想范畴,即模型与数据结合的研究模式<sup>[7]</sup>,体现宏观大气运动规律与局地观测数据的较好结 合,因此该方法在很大程度上优于纯粹的数据挖掘技术(无机理)或纯粹的基于物理规律的技术(强机理).

#### 1.1 观测数据

本文的研究区域分布着 8 个较稳定的地面气象观测站,分别为安多、班戈、比如、嘉黎、林芝、纳木 错、那曲和聂荣站.本论文采用交叉检验(cross-validation)的方法,将 7 个观测站的数据用于训练算法, 剩下 1 个观测站选为测试站.本文采用 2014 年 8 月 8 日 02:00 至 2014 年 8 月 29 日 02:00 期间近地面 层 10 m 逐半小时水平风数据,因此每个观测站共有 312 个连续的观测数据值. 8 个观测站分布在研究 区域内 3 km 等间距分布的 2 160 个网格点之间,具体格点与训练观测站以及测试观测站分布如图 2 所示.本文采用的方法适用于观测站与网格点位置的任意分布情况.



图 2 格点、训练观测站以及测试观测站分布图

### 1.2 WRF 模式

WRF模式是一种中尺度天气预报模式,其先进的数据同化技术、功能强大的嵌套能力和先进的物理 过程使得它适用范围广泛.WRF模式反映大气的物理运动规律.而由于地形和气象现象的复杂性,WRF 模式在区域性和中小尺度的模式输出方面通常具有较高的系统偏差.本文将充分利用观测数据、模式输出 和算法相结合的方式,确保重构数据既满足大气运动规律又同时具有局部测量数据的特征.

WRF 模式计算所用的初始数据为美国国家环境预报中心 2014 年的实时 gft 风数据并且参考了文献 [8]中针对青藏高原设计的 WRF 物理参数.在本文中设置了两层嵌套网格,每层均设置 36 \* 60 个网格点, 内外层网格格距分别为 3 km 和 9 km. WRF 模式输出值 WRFout 为离地面 10 m 高风场纬向分量 $U_{10}$  和径 向分量 $V_{10}$ ,输出值间的时间间隔为 30 min.图 3 显示 WRF 运算所得风速值在大部分时间点处高于实际观 测风速值,因此该现象可理解为 WRF 系统性误差.

# 2 Gappy POD 及其应用

在观测数据严重缺失时经典 POD 方法往往由于无法求解 L<sup>2</sup> 范数从而失效. Gappy POD 的主要思想 为通过建立面具阵(mask matrix),将求解 L<sup>2</sup> 最小值的问题降纬至已有数据的纬度从而将原问题转化为低 纬度上的 L<sup>2</sup> 范数最值问题.

## 2.1 数据分割

文中  $d_1$ , $d_2$ , $d_3$  分别表示网格点数、训练所用站点数和检验测试所用站点数,且相应数值分别为  $d_1 = 2$  160,  $d_2 = 7$ ,  $d_3 = 1$ .

定义 $\tilde{U}, \tilde{V} \in \mathbb{R}^{d \times N}$ (其中 $d = d_1 + d_2 + d_3 = 2$ 168表示空间点, N = 1008表示时间点), 表示 WRF模式计算所得所有空间点(包括网格点和观测站)的水平风速值:





图 3 7个训练观测站点处观测风速和 WRFout 风速的比较

其中:每一列 $u_j, v_j \in \mathbb{R}^d$ ,  $j = 1, \dots, N$ ;  $d = d_1 + d_2 + d_3$ 表示每半个小时的数据集;每一行表示每个空间点在不同时间点的模式水平风场数据.  $\tilde{U}, \tilde{V} \in \mathbb{R}^{d \times N}$ 矩阵的前 $d_1$ 行表示网格点处的风场数据, $d_1 + 1$ 到 $d_1 + d_2$ 行表示 7个训练观测站点处的风场数据,最后一行对应第8个观测站处的风场数据,且该站点位置和数据将用于测试和检验以及评价该方法的有效性.

为了推导方便,将训练集和测试集分别写成如下形式:

 $u_{1N}$  $u_{2i}$  $v_{2i}$  $u_{2N}$  $v_{21}$  $v_{2N}$ U =(2)•••  $u_{d_1+d_2,N}$  $\mathcal{U}_{d_1+d_2}, 1$  $u_{d_1+d_2,j}$  $\mathcal{U}_{d_1+d_2}, j$  $\mathcal{U}_{d_1+d_2}, N$  $(u_{d_{2},N})$ 和 $V = (v_{d_{2},1}, \dots, v_{d_{2},j}, \dots, v_{d_{2},N})$ ,其中(U, V)和(U, V)分别表示训练集  $\Rightarrow U = (u_{d-1})$ , u<sub>d.,i</sub>,  $\begin{pmatrix} m{U} \\ m{U} \end{pmatrix}$ ,  $\widetilde{m{V}}$  = 和测试集,因此 $\tilde{U}$ =

#### 2.2 POD 基的建立

 $\dot{\langle u_j, v_j \rangle} \in \mathbb{R}^{(d_1+d_2)}$  与 $\dot{U}, \dot{V} \in \mathbb{R}^{(d_1+d_2)\times N}$  为样本的中心平均值. 文中为了简化符号,分别用 $u_j, v_j, U, V$ 

代替 $u_j$ , $v_j$ ,U,V 且 $\Lambda^u$ , $\Lambda^v \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$  和 $W^u$ , $W^v$  表示对应协方差矩阵 $U^{\mathsf{T}}U$ , $V^{\mathsf{T}}V \in \mathbb{R}^{(d_1+d_2)\times(d_1+d_2)}$  的特征值和特征向量. 按照特征值的大小顺序对特征向量进行排序,选取前 q 个主特征向量  $W_q^u$ , $W_q^v \in \mathbb{R}^{(d_1+d_2)\times q}$ ,则 POD 基可定义为如下形式:

$$\boldsymbol{U}_{q} = \frac{\boldsymbol{U}_{q} \boldsymbol{W}_{q}^{u}}{\|\boldsymbol{U}^{\mathrm{T}} \boldsymbol{U}\|} \in \mathbb{R}^{(d_{1}+d_{2})\times q} \qquad \boldsymbol{V}_{q} = \frac{\boldsymbol{V}_{q} \boldsymbol{W}_{q}^{v}}{\|\boldsymbol{V}^{\mathrm{T}} \boldsymbol{V}\|} \in \mathbb{R}^{(d_{1}+d_{2})\times q}$$
(3)

假设 $u_i, v_i \in \mathbb{R}^{(d_1+d_2)}, j = 1, 2, \dots, N.$ 为7个训练观测站的观测数据集,则问题转化为求解如下最值问题:

$$\min_{a_j} \| \boldsymbol{U}_q \boldsymbol{a}_j - \overset{\circ}{\boldsymbol{u}}_j \|_{L^2} \qquad \min_{b_j} \| \boldsymbol{V}_q \boldsymbol{b}_j - \overset{\circ}{\boldsymbol{v}}_j \|_{L^2}$$
(4)

由于 $u_j$ , $v_j$  仅代表观测数据,其中对应于 $U_q a_j$ , $V_q b_j$  的格点数据缺失,导致(4)式中各项的数据纬度出现严重差异.

# 2.3 面具阵(mask matrix)的建立

令矩阵  $P \in \mathbb{R}^{d_2 \times (d_1 + d_2)}$  具有如下形式,

$$\boldsymbol{P} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = (\boldsymbol{O} \ \boldsymbol{I})$$
(5)

其中: **O** 为 $d_2 \times d_1$  的零矩阵; **I** 为 $d_2 \times d_2$  的单位阵. 矩阵 **P** 能够使 $d_2 \times (d_1 \times d_2)$  纬度问题降至 $d_2 \times d_2$ 纬度, 即 $d_2 \times d_1$  纬度信息被面具形式隐藏, 因此该 **P** 矩阵取名为面具阵. 将该面具阵作用于原最小值问题, 从而得

$$\min_{a_j} \| \boldsymbol{P}(\boldsymbol{U}_q a_j) - \boldsymbol{P} \boldsymbol{u}_j \|_{L^2} \qquad \min_{b_j} \| \boldsymbol{P}(\boldsymbol{V}_q b_j) - \boldsymbol{P} \boldsymbol{v}_j \|_{L^2}$$
(6)

(6) 式中各项维度均为  $d_2 \times d_2$ ,因此该问题转化为经典最小值求解问题.针对存在缺失数据的 POD 问题, Gappy POD 算法<sup>[9-16]</sup> 可通过面具阵将模型中存在缺失数据的项移除(隐藏).经简易推导可求得上述问题 解的表达式如下:

$$a_{j} = \left[ (\boldsymbol{P}\boldsymbol{U}_{q})^{\mathrm{T}} (\boldsymbol{P}\boldsymbol{U}_{q}) \right]^{-1} (\boldsymbol{P}\boldsymbol{U}_{q})^{\mathrm{T}} (\boldsymbol{P}\boldsymbol{u}_{j}) \qquad b_{j} = \left[ (\boldsymbol{P}\boldsymbol{V}_{q})^{\mathrm{T}} (\boldsymbol{P}\boldsymbol{V}_{q}) \right]^{-1} (\boldsymbol{P}\boldsymbol{V}_{q})^{\mathrm{T}} (\boldsymbol{P}\boldsymbol{v}_{j})$$
(7)

## 2.4 风速重构效果检验和风场重构

本文结合 7 个训练观测站数据以及研究区域内 WRFout 格点数据,通过 Gappy POD 算法重构了研究 区域内任意点处离地 10 m 的水平风场数据,取得了较好的效果.

首先,经比较可知,测试站点处单独采用 WRF 模式计算所得风速平均误差为 2.262 1,而经过 Gappy POD 方法进行风速重构后平均误差可降至 1.838 4(如图 4 所示,其中蓝色曲线表示 WRF 模式在测试站点 处的风速重构误差,红色曲线表示采用 Gappy POD 后在测试站点处风速重构的误差,两条直线分别为对 应颜色曲线的线性拟合),因此当观测数据质量较差时,如某些站点观测数据与真实情况相差较大以及站 点的数据在某些时间段存在缺失等现象时,Gappy POD 方法可用于提升观测质量.

其次, Gappy POD 方法可应用于提取气象要素的宏观特征.图 5 展示了采用 Gappy POD 算法在 26 h 内的 6 个时间点所重构的所有网格点处的水平风场.由于网格点个数远大于观测站点的个数,因此所重构 的每个格点处的数据精度有待进一步提升.然而该方法不同于数据同化,可以在采用较低计算成本的基础 上对气象要素在研究区域内的整体特征提供一定的参考信息.

最后,本文方法可用于重构研究区域内的任意点作为的预测点,均能取得较好的重构效果.比如,在 那曲市偏僻山区处建设信号基站时,可采用本文方法重构候选点处的高精度风速数据.





图 5 6个时间点的水平风场重构图

# 3 结果

本文通过研究西藏那曲一带地形复杂和观测站分布稀疏区域的风速重构问题,介绍了流体力学领域处 理缺失值问题的 Gappy POD 方法,并且将该方法应用于风速重构问题.文中采用 Gappy POD 算法将模式 与观测数据结合的方式提升了模式的重构精度.本文介绍的方法在处理小尺度且数据量小的气象要素重构 问题时具有精度高、易于操作以及计算成本低的特点,并且文中介绍的方法可以应用于解决其他气象要素 的缺失值处理和重构问题.本文也展示了采用该方法重构所有格点处水平风场数据的结果,然而数据同化 并非本文研究的目的和文中方法的主要特点,因此采用该方法重构所有格点数据的结论仅在认识研究区域 内气象要素宏观特征时做为参考.

#### 参考文献:

- [1] WHITEMAN C D. Mountain Meteorology [M]. Oxford: Oxford University Press, 2000.
- [2] PIELKE R A. Mesoscale Meteorological Modeling [M]. Salt Lake: American Academic Press, 2002.
- [3] 谈建国,常远勇,韩志惠,等.一种基于粗糙度分布的大风风速降尺度方法初步研究——以上海市域为例 [J]. 气象科 学,2017,37(6):825-831.
- [4] 杨艳蓉,曾明剑. 雷达资料同化对暴雨预报影响的数值模拟研究 [J]. 气象科学, 2012, 32(2): 145-152.
- [5] 吴佳敏,郑有飞,吴荣军,等. 基于 WRF-3DVAR 同化多源融合数据对近海风模拟的改进试验 [J]. 气象科学,2017, 37(1):120-126.
- [6] 方艳莹,徐海明,朱蓉,等. 基于 WRF 和 CFD 软件结合的风能资源数值模拟试验研究 [J]. 气象,2012,38(11): 1378-1389.
- [7] 邵建,张肃诏,李强,等.智能网格预报时空协调一致关键技术研发[J]. 气象科学,2019,39(6):847-852.
- [8] YANG J H, DUAN K Q, WU J K, et al. Effect of Data Assimilation Using WRF-3DVAR for Heavy Rain Prediction on the Northeastern Edge of the Tibetan Plateau [J]. Advances in Meteorology, 2015, 2015: 294589.
- [9] BUI-THANH T, DAMODARAN M, WILLCOX K. Aerodynamic Data Reconstruction and Inverse Design Using Proper Orthogonal Decomposition [J]. AIAA Journal, 2004, 42(8): 1505-1516.
- [10] LEE K, MAVRIS D N. Unifying Perspective for Gappy Proper Orthogonal Decomposition and Probabilistic Principal Component Analysis [J]. AIAA Journal, 2010, 48(6): 1117-1129.
- [11] MURRAY N, SEINER J. The Effects of Gappy POD on Higher-Order Turbulence Quantities [C] //46th AIAA Aerospace Sciences Meeting and Exhibit. Virginia: AIAA, 2008: 241.
- [12] MURRAY N E, UKEILEY L S. An Application of Gappy POD [J]. Experiments in Fluids, 2007, 42(1): 79-91.
- [13] MIFSUD M, VENDL A, HANSEN L U, et al. Fusing Wind-Tunnel Measurements and CFD Data Using Constrained Gappy Proper Orthogonal Decomposition [J]. Aerospace Science and Technology, 2019, 86: 312-326.
- [14] SCHNEIDER T. Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values [J]. Journal of Climate, 2001, 14(5): 853-871.
- [15] WILLCOX K. Unsteady Flow Sensing and Estimation via the Gappy Proper Orthogonal Decomposition [J]. Computers
  & Fluids, 2006, 35(2): 208-226.
- [16] TSERING-XIAO B, XU Q W. Gappy POD-Based Reconstruction of the Temperature Field in Tibet [J]. Theoretical and Applied Climatology, 2019, 138(1-2): 1179-1188.

#### 责任编辑 张枸