

# 支持向量机估算成都近地面大气光学湍流

吕洁<sup>1,2</sup>, 朱文越<sup>1</sup>, 青春<sup>1,2</sup>, 吴晓庆<sup>1</sup>, 陈莎莎<sup>1,2</sup>

<sup>1</sup>中国科学院安徽光学精密机械研究所大气光学重点实验室, 安徽 合肥 230031;

<sup>2</sup>中国科学技术大学研究生院科学岛分院, 安徽 合肥 230026

**摘要** 介绍了一种估算大气光学湍流的方法——支持向量机。依据该方法用温度梯度、风速和相对湿度估算了 2014 年 5 月 13 日至 2014 年 5 月 18 日四川成都平原地区近地面连续 6 天的大气折射率结构常数  $C_n^2$ , 并将其与实际观测值进行了比较。结果表明两者吻合得较好, 能基本表现成都地区的光学湍流特征。与 Wyngaard 算法相比, 支持向量机得到的结果与观测值更加贴近, 其夜间估算的准确度高于 Wyngaard 算法。相关分析结果表明: 估算值与观测值的相关度在 80% 以上。平均绝对误差与平均相对误差等误差分析显示, 训练好的支持向量机能够基本准确地估算出成都地区的  $C_n^2$ , 但转换时刻的估算值与观测值仍有一定误差。该研究表明, 进一步提高估算精度需要改进模式在转换时刻的估算结果。

**关键词** 大气光学; 折射率结构常数; 支持向量机; 估算模型

中图分类号 P183.4

文献标识码 A

doi: 10.3788/CJL201845.0404001

## Estimation of Atmospheric Optical Turbulence at Near Surface of Chengdu with Support Vector Machine

Lü Jie<sup>1,2</sup>, Zhu Wenyue<sup>1</sup>, Qing Chun<sup>1,2</sup>, Wu Xiaoqing, Chen Shasha<sup>1,2</sup>

<sup>1</sup>Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, Anhui 230031, China;

<sup>2</sup>Science Island Branch of Graduate School, University of Science and Technology of China, Hefei, Anhui 230026, China

**Abstract** A method of estimating atmospheric optical turbulence is introduced which is called support vector machine (SVM). Based on this method, the gradient of temperature, wind speed, and relative humidity are used to calculate the refractive structure parameter  $C_n^2$  in Chengdu, Sichuan from 2014-5-13 to 2014-5-18, which is a continuous 6-day estimated result. The estimated results agree well with the measurement results, which can basically show the characteristics of optical turbulence in Chengdu area. Compared with results of Wyngaard algorithm, the results of SVM are more consistent with the measurements, and the estimation accuracy at night is higher than that of the Wyngaard algorithm. The relative analysis demonstrates that the correlation coefficient between the estimations and observations is over 80%. Besides, mean absolute error and mean relative error prove that after training, SVM can precisely estimate  $C_n^2$  in Chengdu. However, deviations are found between estimations and measurements during the transition time, which should be overcome to improve the performance of the model.

**Key words** atmospheric optics; refractive structure parameter; support vector machine; estimation model

**OCIS codes** 010.1290; 010.1330

## 1 引言

光学湍流是发生在大气中的一种现象, 当它发生时, 会使通过大气传输的光束质量严重下降, 影响发

射和光接收系统的跟瞄精度, 因此了解大气光学湍流的特性是解决这些问题的有效方式。目前, 主要用大气折射率结构常数  $C_n^2$  来衡量光学湍流强度。 $C_n^2$  不是常规的观测参数, 需要搭载专门的平台进行观测, 目

收稿日期: 2017-08-29; 收到修改稿日期: 2017-10-13

作者简介: 吕洁(1991—), 女, 硕士研究生, 主要从事大气光学湍流估算方面的研究。E-mail: lvjie721@mail.ustc.edu.cn

导师简介: 朱文越(1976—), 男, 博士, 研究员, 博士生导师, 主要从事大气光传输的湍流效应方面的研究。

E-mail: zhuwenyue@aiofm.ac.cn(通信联系人)

前在新疆等地已进行了一些观测实验,并给出了大气相干长度和等晕角等测量值<sup>[1]</sup>,但长时间、大范围观测  $C_n^2$  并不现实。因此,寻找一种能够利用容易获得的常规气象参数估算  $C_n^2$  的方法有助于解决这一问题。

国内外研究人员在设计  $C_n^2$  估算模型方面做了大量工作。Tatarski<sup>[2]</sup>于1961年给出了  $C_n^2$  与常规气象参数的关系,即 Tatarski 模式,但缺点是外尺度难以确定;Wyngaard 等<sup>[3]</sup>以 Kansas 平原的实验结果为基础,发现温度结构参数  $C_T^2$  与高度  $z$  之间的关系,据此提出在不稳定层结下估算  $C_T^2$  的经验函数,引入无量纲常数,并依据  $C_T^2$  与  $C_n^2$  的关系提出了估算  $C_n^2$  的公式;通过实验发现这种基于 Monin-Obukhov 相似(MOS)理论估算  $C_n^2$  的方法在不稳定层结下具有较高的准确度<sup>[4-5]</sup>,但在中性和稳定层结下需要对其进行改进<sup>[6]</sup>;Sadot 等<sup>[7]</sup>利用沙漠地区观测的水汽、温度等常规气象参数,通过构建多元非线性回归方程估算  $C_n^2$ ,这是一种简单可行的方式;区域中尺度预报模型能够使物理化参数过程更加客观精细,具有代表性的中尺度气象模式有第 5 代中尺度模式(MM5)和气象数值模式(WRF),这两种模式在国内外均有应用,已有学者借助其预报结果预测  $C_n^2$  高空廓线,这可为天文台的选址提供参考<sup>[8]</sup>。青春等<sup>[9-10]</sup>利用 WRF 模拟了泰山站、高美古等不同下垫面的大气湍流廓线,基本实现了估算  $C_n^2$  廓线的目标。随着大数据时代的到来,机器学习算法兴起,从观测数据(训练样本)出发,在统计学习理论中寻找规律,利用这些规律对未来或无法观测的变量进行预测<sup>[11]</sup>。Wang 等<sup>[12]</sup>使用神经网络在 Mauna Loa 附近海面对  $C_n^2$  进行了为期一个月的估算,初步证实了这种估算  $C_n^2$  方法的可靠性。

本文在机器学习基础上,使用一种机器学习算法:支持向量机(SVM),基于 2014 年 5 月 12 日在四川成都地区观测的数据,对 2014 年 5 月 13 日至 2014 年 5 月 18 日的  $C_n^2$  进行估算,并将估算结果与观测结果进行比较。同时,为评价 SVM 模型的准确度与可靠性,采用平均绝对误差(MAE)和平均相对误差(MRE)等统计量进行描述,初步证实了利用 SVM 方法估算成都地区  $C_n^2$  的可靠性。

## 2 理 论

### 2.1 估算 $C_n^2$ 的理论方法

若湍流局地均匀且各向同性,则折射率结构函数  $D_n(\mathbf{r})$  可定义为

$$D_n(\mathbf{r}) = \langle [n(\mathbf{x}) - n(\mathbf{x} + \mathbf{r})]^2 \rangle, \quad (1)$$

式中  $n(\mathbf{x})$  为折射率,  $\mathbf{r}, \mathbf{x}$  为位置矢量,  $\langle \cdot \rangle$  为系综平均。折射率结构函数  $D_n(\mathbf{r})$  与  $C_n^2$  的关系式为

$$D_n(\mathbf{r}) = C_n^2 r^{2/3}, l_0 \ll r \ll L_0, \quad (2)$$

式中  $l_0$  和  $L_0$  分别为湍流的内尺度和外尺度。由(1)、(2)式可以得到:

$$C_n^2 = \langle [n(\mathbf{x}) - n(\mathbf{x} + \mathbf{r})]^2 \rangle r^{-2/3}. \quad (3)$$

同理,对于温度  $T(\mathbf{x})$ ,温度结构常数  $C_T^2$  可表示为

$$C_T^2 = \langle [T(\mathbf{x}) - T(\mathbf{x} + \mathbf{r})]^2 \rangle r^{-2/3}. \quad (4)$$

对于可见光和近红外波段的波长  $\lambda$ ,折射率结构常数与温度结构常数之间的关系为

$$C_n^2 = 79 \times 10^{-6} \times \frac{P}{T^2(\mathbf{x})} C_T^2, \quad (5)$$

式中  $P$  为气压,单位为 hPa。所使用的测量数据来自温度脉动仪,温度脉动仪通过两个相距为 1 m 的微温探头将空间两点环境温度的变化感应为电阻值的变化,则从电压放大器输出的电压变化  $\Delta V$  与温度变化  $\Delta T$  一一对应,即

$$\Delta V = A \cdot \Delta T, \quad (6)$$

式中  $A$  为标定系数,由(4)式和(6)式可以计算得到温度结构常数  $C_T^2$ ,将其代入(5)式得到折射率结构常数  $C_n^2$ ,这也是温度脉动仪的测量原理。

在目前估算  $C_n^2$  的方法中,基于 MOS 相似理论估算  $C_n^2$  是目前应用较广泛的方法,即只需要获得高度不同的两层的风速和温度,即可估算  $C_n^2$ 。Wyngaard 于 1971 年提出  $C_T^2$  与高度  $z$  满足的关系式为

$$\frac{C_T^2}{z^{4/3} (\partial \bar{\theta} / \partial z)^2} = g_T(R_{ig}), \quad (7)$$

式中  $\bar{\theta}$  为平均位温; $g_T$  为相似函数<sup>[3]</sup>;  $R_{ig}$  为梯度理查森数,用来判定大气层结是否稳定,通常,当  $1/R_{ig} > 4$  时,大气处于不稳定状态。 $R_{ig}$  的表达式为

$$R_{ig} = \frac{g}{\bar{\theta}} \frac{\partial \bar{\theta} / \partial z}{(\partial \bar{u} / \partial z)^2 + (\partial \bar{v} / \partial z)^2}, \quad (8)$$

式中  $\bar{u}$  和  $\bar{v}$  表征速度的水平分量,  $g$  为重力加速度。由(8)式得到梯度理查森数,然后将计算得到的  $R_{ig}$  代入(7)式中得到  $C_T^2$ ,最后代入(5)式计算出  $C_n^2$ ,该方法即为 Wyngaard 于 1971 年提出的估算  $C_n^2$  的方法(下文简称 W71 方法)。

### 2.2 SVM

Vapnik 等基于统计学习理论提出了一种小样本学习机理论——SVM。与传统的机器学习理论不同,SVM 服从结构风险最小化原理而非经验风险

最小化原理,在最小化样本点误差的同时缩小模型泛化误差的上界,从而提高了模型的泛化能力,使模型能够在更大程度上避免维数灾、过学习等问题。

SVM 通过构建一个最优超平面,将各类样本分开,距离最优超平面最近的向量称为支持向量,SVM 可分为支持向量分类机(SVC)和支持向量回归机(SVR),SVR 又可分为线性 SVM 和非线性 SVM。若遇到的问题本质上为非线性问题,则需要引入核函数,将在低维的非线性问题转化为高维的线性回归问题,核函数的引入在某种程度上能够避免维数灾的问题。

给定一组训练集  $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ , SVM 旨在找到一个最优的函数表达式<sup>[13]</sup>  $f(x) = \langle \omega, x \rangle + b$ , 其中  $f(x)$  为估计函数,  $\langle \cdot \rangle$  表示内积运算,  $\omega$  为一组权重系数,  $b$  为偏置值。令  $f(x) = 0$ , 则对应的划分超平面可表示为

$$\langle \omega, x \rangle + b = 0. \quad (9)$$

在 SVM 中,为了寻找最优划分超平面,需要满足所有样本点与最优划分超平面的距离最小,这样所有的样本点都落在超平面的两条边界线之间,而求得的最优函数则只与落在这两条边界线上的支持向量有关,而落在其之外的点对结果没有影响<sup>[14]</sup>。图 1 为由 SVR 确定的最优超平面的粗略图示,图中  $\epsilon$  为损失函数。

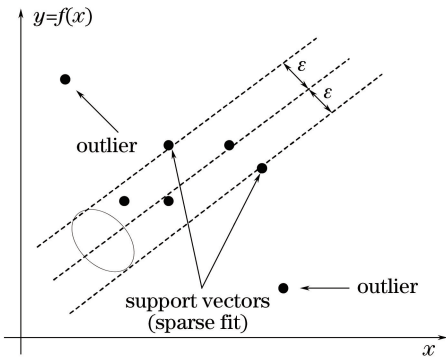


图 1 由 SVR 确定的最优超平面

Fig. 1 Optimal hyperplane determined by SVR

考虑单位平面,最优划分超平面满足<sup>[14]</sup>:

$$\min_i |\langle \omega, x_i \rangle + b| = 1, \quad (10)$$

则在该平面上的所有点满足约束条件:

$$y_i [\langle \omega, x_i \rangle + b] \geq 1, \quad i = 1, 2, \dots, k, \quad (11)$$

式中  $k$  为满足条件的第  $k$  个点。两条边界  $l_1$  和  $l_2$  之间的最大间隔为  $2/\|\omega\|$ , 因此,通过求最大间隔  $\rho$  进而求出最优划分超平面等效于求解二次规划的问题<sup>[15]</sup>:

$$\min_{d, \omega, b} \frac{1}{2} \|\omega\|^2 \quad \text{s.t.} \quad y_i [\langle \omega, x_i \rangle + b] \geq 1, \quad (12)$$

$$i = 1, 2, \dots, l.$$

将根据所选择的误差函数确定最优划分超平面的方法定义为最优方法,即要求通过函数关系  $y = f(x)$  得到的估计值与真实值之间的累积误差最小,这里选择不敏感损失函数。

实际上,部分样本点会落在最优回归超平面之外,无法做到完全线性分离,此时引入松弛变量  $\xi_i$  修改目标函数和约束条件,则出现新的二次凸规划问题,即

$$\min \left[ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \right] \quad \text{s.t.} \quad \begin{cases} y_i - \langle \omega \cdot x_i \rangle - b \leq \epsilon + \xi_i \\ \langle \omega \cdot x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (13)$$

式中  $C$  为惩罚因子<sup>[15]</sup>。由于目标函数是凸函数,因此存在全局唯一最小解,应用 Lagrange 乘子并考虑满足 Karush-Kuhn-Tucker (KKT) 条件,引入核函数,将在低维的非线性问题转化为高维的线性回归问题,最后得到新的回归方程<sup>[16]</sup>:

$$f(x) = \langle \omega, x \rangle + b = \sum_{SV_s} (\alpha_i - \alpha_i^*) K(x, x_i) + b, \quad (14)$$

式中  $K(x, x_i) = \langle \varphi(x), \varphi(x_i) \rangle$  为核函数,  $\varphi(x)$  表示某种非线性变换函数,  $\alpha_i$  和  $\alpha_i^*$  为 Lagrange 乘子,  $SV_s$  表示支持向量。高斯核函数,又称 RBF 核函数,可表示为

$$K'(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right), \quad \sigma > 0, \quad (15)$$

式中  $\sigma$  为高斯核宽。

综上, SVM 首先通过核函数将输入场变换到高维空间,在这个高维空间求最优划分平面(广义),它在形式上类似一个神经网络,最后的输出是中间节点的线性组合,每个中间节点对应一个支持向量<sup>[17]</sup>。图 2 表示 SVM 的基本工作过程。

### 3 成都地区的 $C_n^2$ SVM 优化模型

#### 3.1 实验数据与模型

实验于 2014 年 5 月在成都地区某处开阔的实验场进行,实验装置包括采集器、传感器、远程控制计算机等。所有仪器分两层安装,分别距离地面 1 m 和 3 m,包括 HMP 155 型温湿传感器、CSAT3 型三维超声风速计和温度脉动仪等。通过该实验可

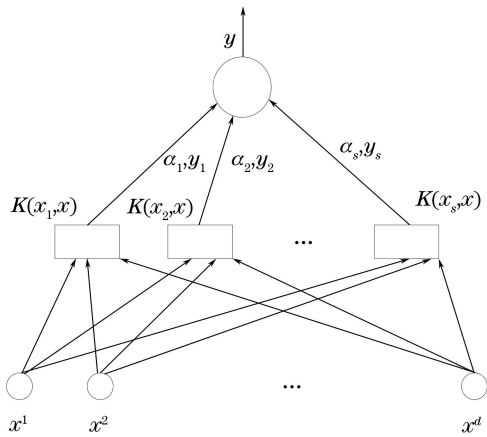


图 2 SVM 的基本工作过程

Fig. 2 Basic working process of SVM

观测三维风速、温度、湿度等常规气象要素。考虑到下雨时温度脉动仪的观测会受到干扰,故选取 2014 年 5 月 12 日至 2014 年 5 月 18 日连续 7 天晴天下完整的观测记录作为实验样本。根据相关性分析,选取 1 m 处的风速、气压和两层温度梯度作为构建  $C_n^2$  估算模型的因子。温度脉动仪的等效  $C_n^2$  分辨率为  $3 \times 10^{-18}$ , 其与 HMP 155 型温湿度传感器和 CSAT3 型三维超声风速计的主要性能参数如表 1 所示,其中  $U_x$ 、 $U_y$  为水平速度分量,  $U_z$  为垂直速度分量, RH 表示相对湿度。

表 1 CSAT3 型三维超声风速计、HMP 155 型温湿度传感器和温度脉动仪的各项性能参数

Table 1 Performance parameters of three-dimensional ultrasonic anemometer CSAT3, temperature humidity sensor HMP 155, and micro-thermal sensor

Instrument name	Measuring parameter	Output range	Offset error	Resolution
CSAT3	$U_x / \text{m} \cdot \text{s}^{-1}$	$\pm 30$	$\pm 0.8$	0.01
	$U_y / \text{m} \cdot \text{s}^{-1}$	$\pm 30$	$\pm 0.8$	0.01
	$U_z / \text{m} \cdot \text{s}^{-1}$	$\pm 8$	$\pm 0.4$	0.005
	$C_n^2 / \text{m}^{-2/3}$			$5 \times 10^{-16}$
HMP 155	RH / %	0.8-100	$\pm 1$	0.1
	Temperature / $^{\circ}\text{C}$	-80-60	$\pm 1.7$	
Micro-thermal sensor	$C_n^2 / \text{m}^{-2/3}$	$10^{-12}$ - $10^{-18}$		$3 \times 10^{-18}$

为了比较建立的估算结果和实际观测结果,将数据分为两部分:第一部分用于建立模型,选取 2014 年 5 月 12 日的 287 条观测数据;第二部分用于模型的检验,选取 2014 年 5 月 13 日至 2014 年 5 月 18 日的 1721 条观测数据。在训练前对数据进行归一化处理,将其转换在  $[0, 1]$  区间,然后取上述

3 个初始预报因子作为训练样本输入场 ( $287 \times 3$  的矩阵);输出为近地面的折射率结构常数  $C_n^2$  ( $1721 \times 1$  的矩阵)。

影响 SVM 模型的关键因素主要包括核函数、损失函数和惩罚常数,即需要确定高斯核宽  $\sigma$ 、损失函数  $\epsilon$  和惩罚常数  $C$ 。如何确定这些参数,目前尚缺乏统一的理论<sup>[18]</sup>。采用循环取值、逐步筛选的办法寻找最合适径向基宽度的步骤为:1)固定其他参数(惩罚因子  $C$  和敏感函数  $e$ ),在一个较大的范围,以较大的步长对径向基参数进行循环取值,当平均绝对误差最小时,确定此时的径向基参数范围;2)缩小步长,重复以上步骤,直至确定出最优的径向基参数。在训练过程中,选取交叉验证误差的方式,使用网格法对算法中的惩罚因子、敏感函数和高斯核宽进行调整。设 3 个参数的取值范围分别为  $C \in \{10^{-2}, 10^{-1}, 1, 10, 10^2\}$ ,  $\sigma \in \{1, 2, \dots, 10\}$ ,  $e \in \{0.1, 0.2, \dots, 1\}$ 。经过第一步的筛选后发现  $C=0.1$ 、 $\sigma=1$ 、 $e=0.1$  时误差最小,接着对以上数据缩小步长。设参数的取值范围为  $C \in \{0.01, 0.02, \dots, 0.2\}$ ,  $\sigma \in \{0.1, 0.2, \dots, 2\}$ ,  $e \in \{0.01, 0.02, \dots, 0.2\}$ 。经过第二轮筛选后最终确定其关键参数为  $\sigma=0.7$ ,  $C=0.16$ ,  $e=0.01$ , 支持向量共 80 个。

将测试样本集、训练好的支持向量机及 Lagrange 乘子  $\alpha_i$  和  $\alpha_i^*$  代入回归方程(14)式中,即可实现对  $C_n^2$  的非线性映射逼近和估算,得到平均相关系数为 84.21%(置信度  $\alpha=0.05$ ),平均绝对误差为 0.2967。

### 3.2 $C_n^2$ 时间序列的观测和估算结果

图 3 显示了 2014 年 5 月 13 日至 2014 年 5 月 18 日连续 6 天的 SVM 的估算结果、观测结果以及 W71 方法的估算结果。实心圆点表示观测结果。结果表明:成都地区的折射率结构常数  $C_n^2$  在 12:00 左右出现一天中的最大值,这可能是因为中午太阳辐射较强,地表辐射增强,湍流能量增强,因此对应的  $C_n^2$  也达到峰值。随后地表吸收的太阳辐射慢慢减弱,对应的湍流能量逐渐减小,因此  $C_n^2$  也开始下降,  $C_n^2$  在一天中出现两个转换时刻,分别约为 06:00 和 18:00。该地区的大气光学湍流强度集中在  $[10^{-16}, 10^{-12}]$  区间内,属于中等偏强的强度。估算结果基本能够准确体现  $C_n^2$  总体的上升、下降和转折趋势,白天的估算结果与观测值符合得最好。图 3 中实心三角形表示 W71 方法的计算结果,与 SVM 的结果相比, W71 方法白天的计算结果与观测结果吻合得较好,但在夜间低风

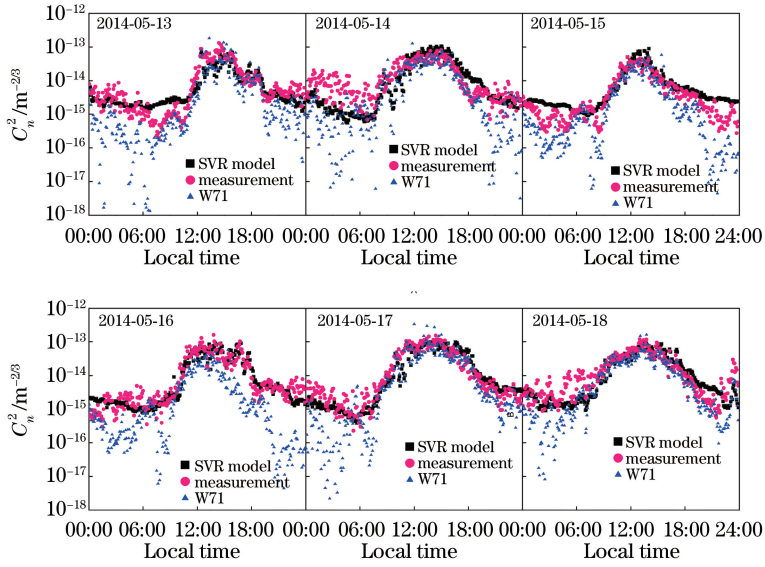


图 3 在成都地区使用 SVM 模型、W71 算法估算  $C_n^2$  的模拟结果和观测结果  
 Fig. 3 Measurement and estimation results of  $C_n^2$  in Chengdu with SVM model and W71 algorithm

速时的计算结果小于观测结果。

### 3.3 SVM 模型结果统计分析

选用 MAE ( $E_{MA}$ )、MRE ( $E_{MR}$ )、均方根误差 (RMSE,  $E_{RMS}$ ) 和相关系数 ( $R_{xy}$ ) 4 个统计量评价 SVM 模型估算的结果,其表达式分别为

$$E_{MA} = \sum_{i=0}^N \frac{|\Delta_i|}{N}, \quad (16)$$

$$E_{MR} = \frac{E_{MA}}{\left| \sum_{i=0}^N \frac{X_i}{N} \right|} \times 100\%, \quad (17)$$

$$E_{RMS} = \sqrt{\sum_{i=0}^N \frac{(\Delta_i)^2}{N}}, \quad (18)$$

$$R_{xy} = \frac{\sum_{i=0}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=0}^N (X_i - \bar{X})^2 \sum_{i=0}^N (Y_i - \bar{Y})^2}}, \quad (19)$$

式中  $X_i$  表示第  $i$  时刻  $C_n^2$  的观测值,  $Y_i$  表示第  $i$  时刻  $C_n^2$  的 SVM 模型估计值,  $\Delta_i = Y_i - X_i$ ,  $N$  表示时间序列的个数,  $\bar{X}$  表示观测值从第 1 个观测时刻到第 1721 个观测时刻的平均值,  $\bar{Y}$  表示对应的估计值的平均值。

图 4 为  $C_n^2$  的观测值和估算值的频数分布直方图和频率分布概率图。图 4(a) 为使用 SVM 模型得到  $\lg C_n^2$  的分析, 由频数分布图可以看出, SVM 模

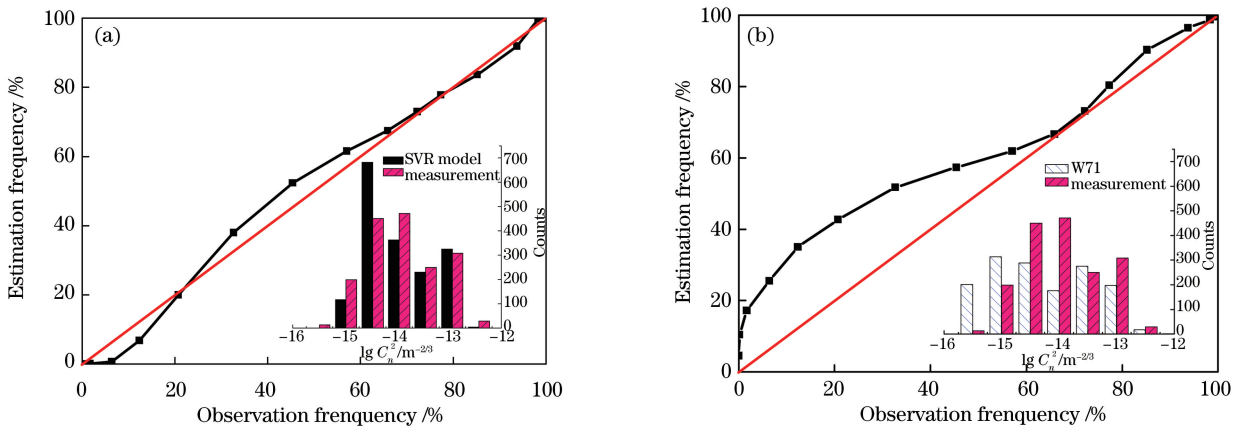


图 4 2014 年 5 月 13 日至 2014 年 5 月 18 日使用不同算法估算  $\lg C_n^2$  得到的频数分布直方图和累积概率分布图。(a) SVM 模型; (b) W71 算法

Fig. 4 Frequency number distribution histogram and cumulative probability distribution of  $\lg C_n^2$  from 2014-05-13 to 2014-05-18. (a) SVM model; (b) W71 algorithm

型的估算结果与观测值分布基本一致,均符合高斯分布。 $[-14.5, -14]$ 区间内 SVM 模型的估算结果的频率大于观测结果;频率概率分布沿直线分布。图 4(b)为 W71 模型得到的估算结果,结果显示在 60%的观测范围内,对应 W71 的频率高于观测值,这也可以从频数分布图中看出,在  $\lg C_n^2$  低于  $-15.5$  时,对应的 W71 估算值频数明显高于观测值,说明 W71 的估算结果偏小,但在  $\lg C_n^2$  大于  $-13.5$  时估算值的频数分布与观测值分布基本一致。

图 5 为  $C_n^2$  的估算值和观测值的散点图。图 5(a)、(b)为 W71 方法在白天(06:00—18:00)的估算结果与夜晚(18:00 至次日 06:00)的估算结果

分别与对应观测值的散点图。图 5(c)、(d)为由 SVM 估算出的估算值分别与对应观测值的散点图。从图 5(a)~(c)可以看出,白天两种模型的点均沿直线分布,相关系数分别为 82.38% 和 88.21%。SVM 模型的估算结果对应的相关度更高,但两者相差不大。在夜晚,W71 对应的估算结果略偏离直线,点偏向观测结果,这是由 W71 方法在夜间的估算结果偏低造成的。SVM 对应的估算结果沿直线分布,两者的相关系数分别为 48.83% 和 66.38%,其他表示误差的参数显示两者差距较大。夜晚 SVM 的估算值明显更贴近观测值,表明 SVM 模型在夜晚仍具有较好的估算能力,尽管不如白天的估算结果准确。

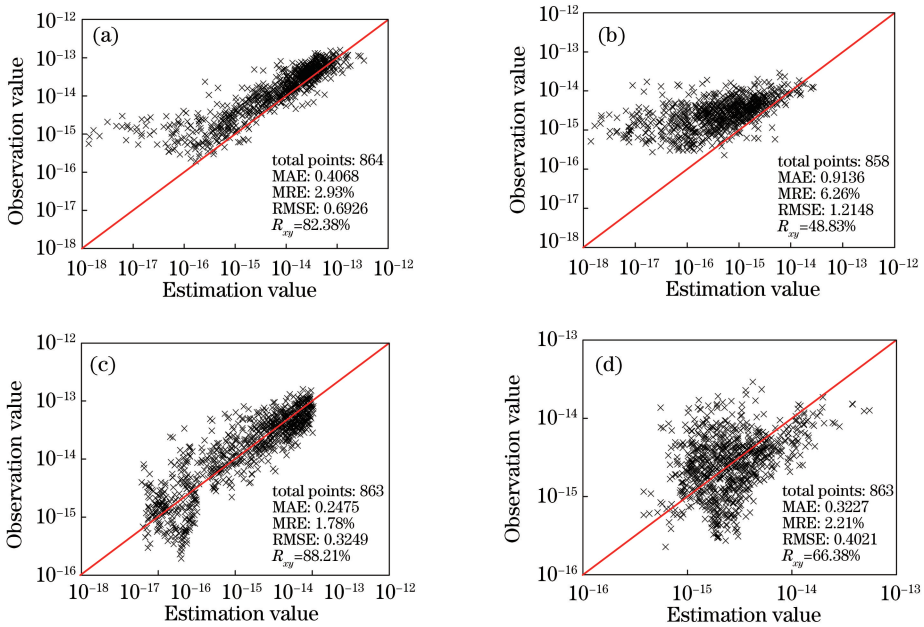


图 5 不同方法得到的  $C_n^2$  估算值和观测值的散点图。(a) W71 算法(06:00—18:00); (b) W71 算法(18:00 至次日 06:00);(c) SVM 算法(06:00—18:00);(d) SVM 算法(18:00 至次日 06:00)

Fig. 5 Scatter plots of estimated and measured values of  $C_n^2$  with different methods. (a) W71 algorithm (06:00—18:00); (b) W71 algorithm (from 18:00 to 06:00 of following day); (c) SVM model (06:00—18:00); (d) SVM algorithm (from 18:00 to 06:00 of following day)

## 4 结 论

使用 SVM 估算了成都地区  $C_n^2$  的时间序列,将其与实测结果进行比较,发现在对成都地区连续 6 天  $C_n^2$  的估算中,SVM 基本能准确估算  $C_n^2$  的变化趋势,总体相关度在 80% 以上,相对误差率在 2.5% 以下。SVM 白天的估算结果相对较好,且与 W71 方法相比在夜晚的估算准确度有所提高。因此,利用 SVM 对  $C_n^2$  进行估算是基本可行的,估算结果能展现出成都地区大气湍流的基本特征。

需要指出的是,SVM 模型是建立在统计理论之

上,误差的累加传递并不明显,即使是最后一天(2014 年 5 月 18 日),其模拟结果依然能够达到较高的准确度。然而,SVM 对成都地区的估算在夜晚的准确度比白天的低,且转换时刻的模拟与观测值仍有差距,这可能与进入模式的因子和使用的核函数参数有关,具体原因需要进一步的研究。同时, $C_n^2$  的估算是一个受制于多种因素的复杂过程,初始因子的挑选是影响实验结果的关键因素,因此更充分地挑选出这些因子将有助于进一步提高  $C_n^2$  的估算准确度,同时 SVM 模型的参数选取目前也缺乏有效的方法,这一问题的解决将在很大程度上显示

SVM 在该领域的优势,这也将是从事光学湍流估算工作者未来可以进一步开展的工作。

参 考 文 献

- [1] Chen X W, Li X B, Liu Q, *et al.* Measurement and analysis of integral optical turbulence at western typical zones[J]. *Acta Optica Sinica*, 2016, 36(8): 0801003.  
陈小威, 李学彬, 刘庆, 等. 中国西部典型地区整层光学湍流观测与分析[J]. *光学学报*, 2016, 36(8): 0801003.
- [2] Tatarski V I. The propagation in a turbulent medium [M]. New York: McGraw-Hill, 1961: 40-58.
- [3] Wyngaard J C, Izumi Y, Collins S A, *et al.* Behavior of the refractive-index-structure parameter near the ground [J]. *Journal of the Optical Society of America*, 1971, 61(12): 1646-1650.
- [4] Andreas E L. Estimating  $C_n^2$  over snow and sea ice from meteorological data [J]. *Journal of the Optical Society of America*, 1988, 5(4): 481-495.
- [5] Davidson K L, Schacher G E, Fairall C W, *et al.* Verification of the bulk method for calculating overwater optical turbulence [J]. *Applied Optics*, 1981, 20(17): 2919-2923.
- [6] Frederickson P A, Davidson K L, Zeisse C R, *et al.* Estimating the refractive index structure parameter ( $C_n^2$ ) over the ocean using bulk methods [J]. *Journal of Applied Meteorology*, 2000, 39(10): 1770-1783.
- [7] Sadot D, Kopeika N S. Forecasting optical turbulence strength on the basis of macroscale meteorology and aerosols: models and validation [J]. *Optical Engineering*, 1992, 31(2): 200-212.
- [8] Wu X Q. Current situation and prospect to mesoscale meteorological model in atmospheric optical turbulence forecast [J]. *Laser & Optoelectronics Progress*, 2017, 54(1): 010001.  
吴晓庆. 中尺度气象模式预报大气光学湍流的现状与展望 [J]. *激光与光电子学进展*, 2017, 54(1): 010001.
- [9] Qing C, Wu X Q, Li X B, *et al.* Estimation of atmospheric optical turbulence profile by WRF model at Gaomeigu [J]. *Chinese Journal of Lasers*, 2015, 42(9): 0913001.  
青春, 吴晓庆, 李学彬, 等. WRF 模式估算丽江高美古大气光学湍流廓线 [J]. *中国激光*, 2015, 42(9): 0913001.
- [10] Qing C, Wu X Q, Li X B, *et al.* Research on simulating atmospheric optical turbulence in typical area [J]. *Acta Optica Sinica*, 2016, 36(5): 0501001.  
青春, 吴晓庆, 李学彬, 等. 典型地区高空大气光学湍流模拟研究 [J]. *光学学报*, 2016, 36(5): 0501001.
- [11] Tian Y J. Support vector regression and its application [D]. Beijing: China Agricultural University, 2005.  
田英杰. 支持向量回归机及其应用研究 [D]. 北京: 中国农业大学, 2005.
- [12] Wang Y, Basu S. Using an artificial neural network approach to estimate surface-layer optical turbulence at Mauna Loa, Hawaii [J]. *Optics Letters*, 2016, 41(10): 2334-2337.
- [13] Chen Y Y, Yu X D, Gao X H, *et al.* A new method for non-linear classify and non-linear regression I: introduction to support vector machine [J]. *Journal of Applied Meteorological Science*, 2004, 15(3): 345-354.  
陈永义, 俞小鼎, 高学浩, 等. 处理非线性分类和回归问题的一种新方法 (I)——支持向量机方法简介 [J]. *应用气象学报*, 2004, 15(3): 345-354.
- [14] Gunn S R. Support vector machine for classification and regression [R]. Southampton: University of Southampton, 1998. [http://read.pudn.com/downloads162/ebook/739616/SVM% 20for% 20% 20classification%20and%20regression.pdf](http://read.pudn.com/downloads162/ebook/739616/SVM%20for%20%20classification%20and%20regression.pdf).
- [15] Smola A J, Scholkopf B. A tutorial on support vector regression [J]. *Statistics and Computing*, 2004, 14(3): 199-222.
- [16] Zeng S H. The theory research of algorithm on support vector regression and application [D]. Chongqing: Chongqing University, 2006.  
曾绍华. 支持向量回归机算法理论研究与应用 [D]. 重庆: 重庆大学, 2006.
- [17] Zhang X G. Introduction to statistical learning theory and support vector machines [J]. *Acta Automatic Sinica*, 2000, 26(1): 32-42.  
张学工. 关于统计学习理论与支持向量机 [J]. *自动化学报*, 2000, 26(1): 32-42.
- [18] Wang D C, Fang T J, Gao L F, *et al.* Support vector machine regression on-line modeling and its application [J]. *Control and Decision*, 2003, 18(1): 89-95.  
王定成, 方廷健, 高理富, 等. 支持向量机回归在线建模及应用 [J]. *控制与决策*, 2003, 18(1): 89-95.