

## 基于支持向量机估算大气光学湍流廓线的研究

朱黎明<sup>1,2\*</sup>, 孙刚<sup>2\*\*</sup>, 陈多龙<sup>2,3</sup>, 张汉九<sup>2,3</sup>, 方圆<sup>2,3</sup>, 马学宾<sup>2</sup><sup>1</sup>安徽大学物质科学与信息技术研究院, 安徽 合肥 230031;<sup>2</sup>中国科学院合肥物质科学研究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>3</sup>中国科学技术大学环境科学与光电技术学院, 安徽 合肥 230031

**摘要** 主要利用支持向量机的机器学习方法开展估算大气光学湍流廓线的研究。通过选取沿海地区实测探空数据, 利用温度、压强、相对湿度、风速、风速切变和温度切变实测廓线数据, 依据支持向量机估算得到不同日期的大气光学湍流廓线, 并与实际测量值进行比较。误差分析结果表明: 在 2018-05-05 和 2018-05-10, 估算的大气光学湍流廓线与实际测量廓线的均方根误差分别为 0.4461 和 0.3939, 相关性分别为 70.42% 和 62.17%。研究证明: 根据实测数据训练学习得到的支持向量机模型能够较为准确地估算沿海地区大气光学湍流廓线, 虽有一定误差, 但大致趋势吻合, 验证了支持向量机方法估算大气光学湍流廓线的可行性, 从而为利用常规气象探空数据直接估算大气光学湍流廓线, 并为建立相关模式的可能性打下基础。

**关键词** 大气光学; 湍流廓线; 支持向量机; 机器学习

中图分类号 P183.4

文献标志码 A

doi: 10.3788/AOS202242.0101001

## Atmospheric Optical Turbulence Profile Estimation Using Support Vector Machine

Zhu Liming<sup>1,2\*</sup>, Sun Gang<sup>2\*\*</sup>, Chen Duolong<sup>2,3</sup>, Zhang Hanjiu<sup>2,3</sup>, Fang Yuan<sup>2,3</sup>, Ma Xuebin<sup>2</sup><sup>1</sup> School of Physics and Materials Sciences, Anhui University, Hefei, Anhui 230031, China;<sup>2</sup> Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, Anhui 230031, China;<sup>3</sup> School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei, Anhui 230031, China

**Abstract** A support vector machine-based machine learning method is proposed to estimate atmospheric optical turbulence profiles. Using sounding data collected in coastal areas, the measured temperature, pressure, relative humidity, wind speed, wind speed shear, and temperature shear profile data are used to estimate the atmospheric optical turbulence profiles on different days. The estimated profiles are compared with the actual measured values. Error analysis results show that the root mean square errors of the estimated atmospheric optical turbulence profile and actual measurement profile are 0.4461 and 0.3939 and the correlation values are 70.42% and 62.17% on 2018-05-05 and 2018-05-10, respectively. The results demonstrate that a support vector machine model trained and learned using actual measured data can accurately estimate the atmospheric optical turbulence profile in coastal areas. Despite some errors, the general trend is consistent. The feasibility of estimating atmospheric optical turbulence profile by support vector machine method is verified, which lays a foundation for directly estimating atmospheric optical turbulence profile by using conventional meteorological sounding data and establishing relevant models.

**Key words** atmospheric optics; turbulence profile; support vector machine; machine learning

收稿日期: 2021-05-08; 修回日期: 2021-06-15; 录用日期: 2021-07-14

基金项目: 中国科学院科技创新重点实验室基金(CXJJ-19S028)

通信作者: \*zlm1998@mail.ustc.edu.cn; \*\*gsun@aiofm.ac.cn

## 1 引言

激光在大气中传输的湍流问题中,主要研究湍流造成的折射率随机变化的规律。大气湍流效应会引起激光光束的漂移、扩展和光强的起伏<sup>[1]</sup>。表征大气光学湍流强度的参数为大气折射率结构常数  $C_n^2$ ,其随高度的分布称为大气光学湍流廓线。国内外许多学者对  $C_n^2$  廓线的估算模型进行了长期的研究和观测,提出了以 SLC 模式<sup>[2]</sup>、AFGL-AMOS 模式<sup>[3]</sup>等为代表的经验模式和以 Hufnagel 模式<sup>[4]</sup>、NOAA 模式<sup>[5]</sup>、Tartaski 模式<sup>[6]</sup>及 Coulman 等<sup>[7]</sup>和 Dewan 等<sup>[8]</sup>提出的各种经验外尺度为代表的参数模式。国内方面,韩亚娟等于 2020 年用 Hufnagel 模式拟合出拉萨地区早晨和夜晚的  $C_n^2$  廓线<sup>[9]</sup>。随着计算机性能的提升,机器学习成为了处理分析数据的新方法、新手段。Wang 等<sup>[10]</sup>于 2016 年利用神经网络对夏威夷 Mauna Loa 附近海面的  $C_n^2$  进行了一个月的预报。吕洁等<sup>[11]</sup>于 2017 年用 BP 神经网络和逐步回归预报三亚连续 4 天的近地面  $C_n^2$  的变化情况。陈小威等<sup>[12]</sup>于 2020 年用遗传算法优化神经网络隐含层单元,得出输入层、隐含层、输出层神经元个数分别为 10、40、1 的 10-40-1 网络架构的多层感知机模型,估算了西北高原地区近地面一个月  $C_n^2$  变化情况。现有机器学习模型主要对近地面  $C_n^2$  进行估算,在复杂探空条件下关于  $C_n^2$  廓线估算和预报的研究还较少。

支持向量机(SVM)是一种新的机器学习方法<sup>[13-15]</sup>,基础是 Vapnik 创建的统计学习理论。与传统神经网络相比,支持向量机针对小样本问题提出,将小样本问题转换为二次规划问题,可以在有限样本的情况下获得最优解。支持向量机不仅可以解决分类、模式识别的问题,还可以解决回归、拟合等问题。因此,支持向量机在各个领域都受到了非常广泛的应用。

本文基于支持向量机方法,利用沿海地区 2018-04-15 至 2018-05-02 气象条件较好的实测探空数据作为训练学习样本,结合 2018-05-05 和 2018-05-10 的常规探空数据对相应时间段的  $C_n^2$  廓线进行估算,并与实测结果进行比较。采用平均绝对误差、均方根误差、平均偏差及相关系数 4 个评价指标进行描述,验证了支持向量机方法在估算大气光学湍流廓线方面的可行性。理论上支持向量机方法具有利用常规气象探空数据直接估算  $C_n^2$  廓线的可能性,为将来大气光学湍流模式的建立和应用提

供了有益的参考。

## 2 实验概述和理论分析

### 2.1 实验概述

2018-04-15 至 2018-05-10,利用探空气球搭载新型 GPS 温度脉动仪在沿海地区开展探空实验。温度脉动仪的统计平均时间为 5 s,频率响应范围为 0.1~30 Hz,温差标准差不大于 0.002 °C,  $C_n^2$  分辨率度为  $2 \times 10^{-18} \text{ m}^{-2/3}$ <sup>[16]</sup>。探空气球的上升速度大约为 5 m/s,垂直分辨率约为 30 m。由于信号传输和天气干扰等因素,尽可能减小不同时间段太阳辐射对测量的影响,均选取上午时段的探空数据进行分析研究,在上午的 23 份的探空资料中选取 15 份质量较好的探空数据,经过预处理后以 100 m 为一层进行平均。获得了大气折射率结构常数  $C_n^2$ 、温度、压强、相对湿度、风速等参数的廓线分布,具体探空记录如表 1 所示。

表 1 沿海地区探空记录

Table 1 Record of balloon soundings in coastland

Balloon number	Launch data	Launch time (local time)	Termination altitude (above sea level) / m
1#	2018. 04. 15	07:43	30360
2#	2018. 04. 16	07:38	32060
3#	2018. 04. 18	07:33	30100
4#	2018. 04. 19	07:33	30490
5#	2018. 04. 20	07:36	29610
6#	2018. 04. 21	07:38	22090
7#	2018. 04. 22	07:37	28000
8#	2018. 04. 25	07:42	25520
9#	2018. 04. 27	07:40	28920
10#	2018. 04. 28	07:38	28620
11#	2018. 04. 30	07:35	30010
12#	2018. 05. 01	07:41	30890
13#	2018. 05. 02	07:39	30490
14#	2018. 05. 05	07:37	31060
15#	2018. 05. 10	07:37	26010

### 2.2 理论分析

对于充分发展的大气湍流,在惯性子区条件下,假设符合 Kolmogorov 的局地均匀各向同性<sup>[17]</sup>,则温度结构函数  $D_T(r)$ 和  $D_n(r)$ 可定义为

$$D_T(r) = \overline{[T(r_1+r) - T(r_1)]^2}, \quad (1)$$

$$D_n(r) = \overline{[n(r_1+r) - n(r_1)]^2}, \quad (2)$$

式中:  $l_0 \leq r \leq L_0$ ,  $l_0$  是湍流的内尺度,  $L_0$  是湍流的外尺度。在可见光和近红外波段,温度变化是影响折射率结构变化的主要因素。因此近似认为折射率结构常数  $C_n^2$  可以直接由某一高度的气压  $P$ , 气温

$T$  和  $C_T^2$  计算所得<sup>[6]</sup>, 即

$$C_n^2(h) = \left[ 79 \times 10^{-6} \times \frac{P(h)}{T(h)^2} \right]^2 \times C_T^2(h). \quad (3)$$

温差变化由温度脉动仪中相距为 1 m 的微温探头获得。本实验中, 微温探头为铂丝, 电阻约 10  $\Omega$ , 直径约为 10  $\mu\text{m}$ 。通过两点的微温探头, 测量空间两点的温差,

$$C_T^2 = \frac{(T_1 - T_2)^2}{r^{2/3}}. \quad (4)$$

将相应的探空气象参数和(4)式所求的  $C_T^2$  代入(3)式, 可得  $C_n^2$  廓线。

大气湍流具有复杂的机制, 湍流强度也与大气层结稳定性密切相关, 大气层结的稳定度常用理查森数来表示, 表达式为

$$R_i = \frac{\frac{g}{\theta} \left( -\frac{\partial \theta}{\partial Z} \right)}{\left[ \left( \frac{\partial U}{\partial Z} \right)^2 + \left( \frac{\partial V}{\partial Z} \right)^2 \right]}, \quad (5)$$

式中:  $\theta$  表示位温;  $\frac{\partial \theta}{\partial Z}$  是位温切变;  $\frac{\partial U}{\partial Z}$  和  $\frac{\partial V}{\partial Z}$  均表示水平风速切变;  $U$  和  $V$  代表水平的两个风速分量。当理查森数小于 0.25 时, 风切变产生的机械湍流会为湍流场源源不断提供能量, 使湍流运动持续发展。

### 2.3 支持向量机

SVM 是 Vapnik 基于统计学习理论提出的一种机器学习算法, 通过寻求最小结构化风险来提高学习的泛化能力, 实现经验风险和置信范围的最小化, 从而达到即使在统计样本量较少的情况下, 也能获得良好统计规律的目的。

SVM 在寻求一个最优超平面时对不同样本进行分类。根据分类和回归问题 SVM 分为支持向量分类机(SVC)和支持向量回归机(SVR)<sup>[13-15]</sup>。给定训练样本集  $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$ , 在分类问题中, 基本想法是基于训练集  $D$ , 在样本空间中找到一个划分超平面, 将不同类别样本分开, 如图 1 所示。

在样本空间中, 划分超平面通过线性方程描述:

$$\boldsymbol{\omega}^T \cdot \boldsymbol{x} + b = 0, \quad (6)$$

式中: 法向量  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)$ ;  $b$  为位移项。超平面参数  $\boldsymbol{\omega}$  和  $b$  一旦确定, 超平面即可确定。对硬间隔且线性可分的分类器, 通过间隔最大化思想, 将直接求解  $\boldsymbol{\omega}$  和  $b$  的问题转换为二次凸优化问题:

$$\min_{\boldsymbol{\omega}, b} \frac{1}{2} \|\boldsymbol{\omega}\|^2 \quad \text{s. t.} \quad y_i(\boldsymbol{\omega}^T \cdot \boldsymbol{x}_i + b) \geq 0,$$

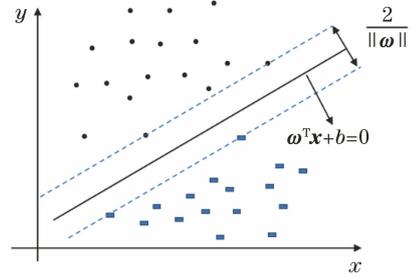


图 1 SVM 确定最优超平面

Fig. 1 Optimal hyperplane determined by SVM

$$i = 1, 2, 3, \dots, N. \quad (7)$$

将(7)式作为原始最优化问题, 构建拉格朗日函数:

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 - \sum_{i=1}^N \alpha_i y_i (\boldsymbol{\omega} \cdot \boldsymbol{x}_i + b) + \sum_{i=1}^N \alpha_i, \quad (8)$$

式中:  $\alpha_i \geq 0$ 。利用拉格朗日函数的满足 Karush-Kuhn-Tucker(KKT)条件的强对偶关系, 通过求解对偶问题得到原始问题的最优解。对偶问题为

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) - \sum_{i=1}^N \alpha_i \\ \text{s. t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \quad y_i(\boldsymbol{\omega}^T \cdot \boldsymbol{x}_i + b) \geq 0, \\ & i = 1, 2, 3, \dots, N. \end{aligned} \quad (9)$$

利用(9)式求得最优解  $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ , 代入

$$\begin{cases} \boldsymbol{\omega}^* = \sum_{i=1}^N \alpha_i^* y_i \boldsymbol{x}_i \\ b^* = y_j - \sum_{i=1}^N \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) \end{cases}, \quad (10)$$

得到  $\boldsymbol{\omega}^*$  和  $b^*$  值, 获得超平面方程, 实现分类决策。为了利用 SVM 解决回归拟合方面的问题, Vapnik 等在 SVM 分类的基础上引入了不敏感损失函数  $\epsilon$ , 得到了支持向量回归机, 如图 2 所示, 基本思想是寻找一个最优分类面, 使所有训练样本离该最优分类

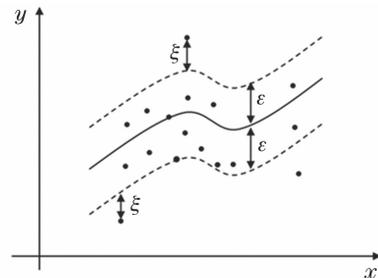


图 2 SVR 确定最优超平面

Fig. 2 Optimal hyperplane determined by SVR

面的误差最小。

类似 SVM 分类,SVR 引入松弛变量  $\xi_i$  和  $\xi_i^*$ , 将上述情况转换为凸优化问题:

$$\begin{aligned} \min & \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s. t. } & y_i - \boldsymbol{\omega}\Phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i, \\ & -y_i + \boldsymbol{\omega}\Phi(\mathbf{x}_i) + b \leq \varepsilon + \xi_i^*, \\ & \xi_i \geq 0, \xi_i^* \geq 0, i=1,2,\dots,N, \end{aligned} \quad (11)$$

式中:最小化函数包含两层含义, $\frac{1}{2} \|\boldsymbol{\omega}\|^2$  代表间隔, $\sum_{i=1}^N (\xi_i + \xi_i^*)$  表征回归的训练误差,应保证间隔尽可能大和训练误差尽可能小; $C$  为惩罚因子,是调和间隔和训练误差的系数; $\Phi(\cdot)$  为转换成对偶形式的符号。同样引入拉格朗日函数,转换为对偶形式:

$$\begin{aligned} \max & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) - \\ & \sum_{i=1}^N (\alpha_i + \alpha_i^*)\varepsilon + \sum_{i=1}^N (\alpha_i - \alpha_i^*)y_i \\ \text{s. t. } & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C, \end{aligned} \quad (12)$$

式中: $K(\mathbf{x}_i, \mathbf{x}_j)$  为核函数,将低维非线性变换的输入空间对应一个高维特征空间中的超平面模型,这样就可以将非线性的问题转换为特征空间中求解线性支持向量机的问题,核函数直接求内积的思想也避免了维数增加造成计算量指数倍增长的维数灾难。求解得出最后的回归函数为

$$f(\mathbf{x}) = \boldsymbol{\omega}^* \Phi(\mathbf{x}) + b^* = \sum_{i=1}^N (\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) + b^*, \quad (13)$$

式中:部分参数 $(\alpha_i - \alpha_i^*)$ 不为零,对应样本  $\mathbf{x}_i$  为支持向量。核函数采用高斯核函数,也称径向基函数(RBF),表示为

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}\right), \sigma > 0, \quad (14)$$

式中: $\sigma$  为高斯核宽,参数调节中用  $g$  表示。综上,SVR 最终的函数形式与 SVM 相同,其结构与神经网络结构相似。图 3 为 SVM 的工作机制示意图,输入层  $\mathbf{x}(i)$  为输入样本,输出是若干中间层节点的线性组合,中间节点对应于输入样本与一个支持向量的内积  $K(\mathbf{x}, \mathbf{x}_n)$ ,因此也被称为支持向量网络。

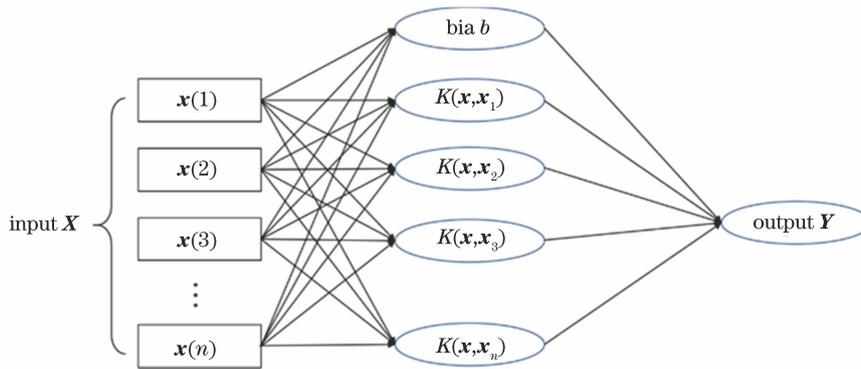


图 3 SVM 工作机制示意图  
Fig. 3 Basic working mechanism of SVM

### 3 估算 $C_n^2$ 的 SVM 模型

#### 3.1 SVM 估算模型建立

$C_n^2$  的变化是多种参数共同作用的复杂过程。结合(3)、(5)式可以推断: $C_n^2$  不仅与常规气象参数有关,还与风速切变和温度切变表征大气层结稳定性的物理量有关。综合考虑  $C_n^2$  的物理机制和模型准确性,选择 6 个常规气象参数作为模型输入,分别为温度  $T$ 、压强  $P$ 、相对湿度  $R_h$ 、风速  $V$ 、温度切变  $\frac{\partial T}{\partial Z}$  和风速切变  $\frac{\partial V}{\partial Z}$ ;选择  $C_n^2$  值为唯一输出构建 SVM 估算模型。

共利用 15 次探空数据,为保证模型的泛化能力,将其中 13 次探空数据作为训练集,用于构建 SVM 估算模型,分别为 2018-04-15、2018-04-16、2018-04-18、2018-04-19、2018-04-20、2018-04-21、2018-04-22、2018-04-25、2018-04-27、2018-04-28、2018-04-30、2018-05-01、2018-05-02 的探空实验数据,共 3986 条记录。其余 2 次探空数据作为测试集代入训练得到的 SVM 估算模型,对  $C_n^2$  进行估算和对比分析,分别是 2018-05-05 和 2018-05-10 的探空实验数据,两次记录数据分别为 310 条和 260 条。图 4 为探空实测的  $C_n^2$  廓线,均为上午 7:30 左右释

放探空气球得到,时间跨度达近一个月,测量高度大多在 20 km 以上,lg C<sub>n</sub><sup>2</sup> 的量级在 -12~-18 之间。

以 2018-05-05 为例,图 5 为 SVM 模型估算 C<sub>n</sub><sup>2</sup> 的输入参数廓线。

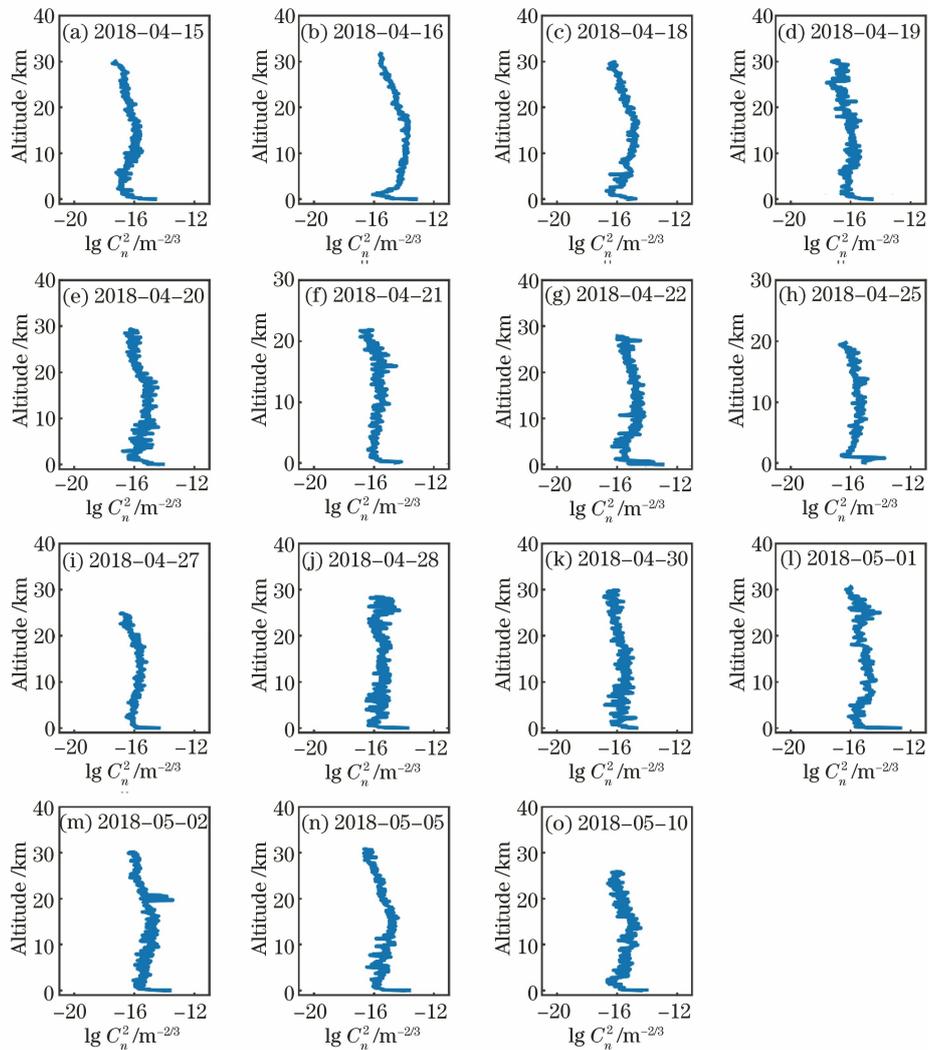


图 4 探空实测廓线

Fig. 4 Profiles of radiosonde measurement

在模型进行训练之前,需要确定一组参数值,称为最优超参数,它可以提高学习的性能和效果,并非通过训练得到。最优超参数的选取需要一定的性能度量来评估,本文中性能度量采用 k 折交叉验证,其思想是将训练集平均分成 N 等份,让每一部分作为一次验证集进行回归,剩下的 N-1 部分再当作训练集,最后所有验证集的平均均方误差当作超参数选取时的性能评价。训练 SVM 模型涉及的三个超参数:惩罚参数 C、核函数参数 g 和损失函数 ε,这里采用网格搜索的穷尽遍历算法,将惩罚参数 C、核函数参数 g 在查找范围 [2<sup>-5</sup>, 2<sup>5</sup>] 进行搜索,设置步长后,以 C 和 g 搜索取值分别为 2<sup>-5</sup>, 2<sup>-4</sup>, …, 2<sup>5</sup> 建立网格,在网格中确定合适的取值范围。如图 6 所示,为了网格坐标线性可视化,横纵坐标分别取以 2

为底的 C 和 g 的对数,等值线为在 C 和 g 网格中搜索确定交叉验证平均均方误差相同点的集合。矩形区域表示在搜索范围中交叉验证均方误差尽可能小的区域,对矩形区域包含的点以 0.25 分辨率进行逐个调试,得到对应的 C=10, g=4。控制 C、g 不变,在 {0.0001, 0.0002, …, 0.01} 范围内调试损失函数 ε,让交叉验证均方误差尽可能小。综上,最后确定 SVM 估算模型的三组超参数的值为 C=10, g=4, ε=0.001,代入模型进行训练,共有 3953 个支持向量。

将 2018-05-05 和 2018-05-10 的常规探空数据作为输入代入训练好的 SVM 估算模型,再将搜索的最优超参数代入,结合回归方程(13)式,即可实现 C<sub>n</sub><sup>2</sup> 廓线的非线性回归。

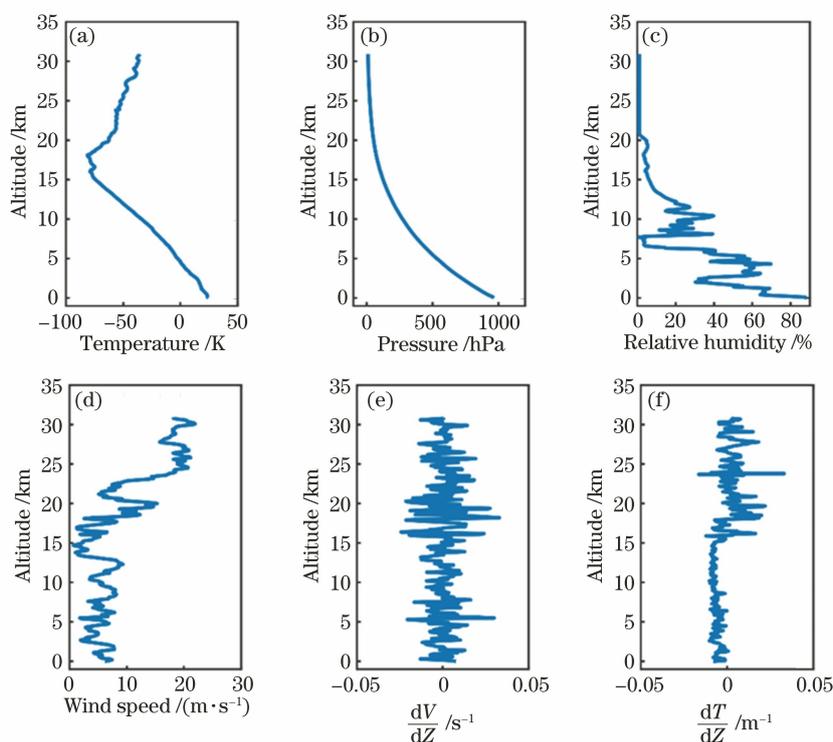


图 5 2018-05-05 常规探空气象参数廓线

Fig. 5 Profiles of conventional radiosonde meteorological parameters on 2018-05-05

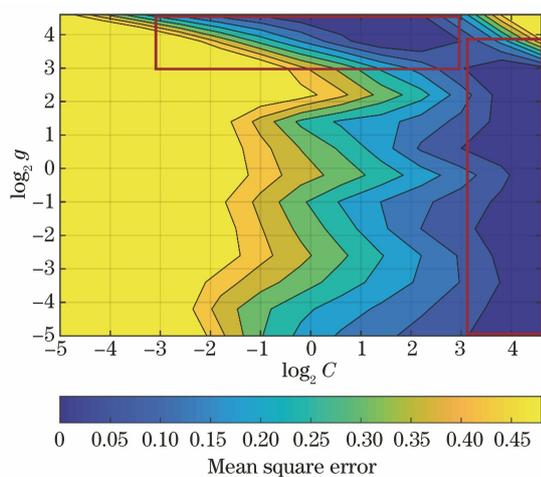


图 6  $C$  和  $g$  的平均均方差等值线

Fig. 6 Contour plot of mean square error of  $C$  and  $g$

### 3.2 SVM 模型估算廓线与实测结果对比

图 7 是 2018-05-05 和 2018-05-10 两天  $C_n^2$  廓线的 SVM 模型估算结果和探空测量结果对比图。结果表明:从趋势上来说,沿海地区的  $\lg C_n^2$  廓线取值范围在  $-18$  至  $-14$  之间,整层变化平缓。SVM 模型估算的  $C_n^2$  廓线能比较准确反映出当日实际测量的  $C_n^2$  廓线分层变化趋势:总体上  $C_n^2$  强度随高度升高而减弱,但在边界层顶大约  $13\text{ km}$  出现强湍流层,即明显的边界层指数递减和对流层增强现象。

尤其是矩形框区域很好地展现了  $C_n^2$  强弱转换的趋势关系。定量上来说,2018-05-05 估算廓线在  $15\text{ km}$  至  $25\text{ km}$  的平流层湍流出现了最多一个数量级的低估,2018-05-10 估算廓线在量级和趋势上均吻合较好。

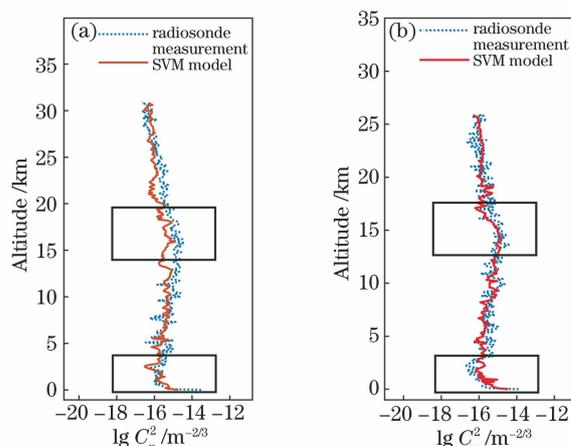


图 7 探空测量和 SVM 模型估算的  $C_n^2$  廓线比较。

(a) 2018-05-05; (b) 2018-05-10

Fig. 7 Comparison of  $C_n^2$  profiles obtained by radiosonde measurement and SVM model. (a) 2018-05-05; (b) 2018-05-10

### 3.3 SVM 模型估算误差分析

为定量描述 SVM 模型的估算效果,选取 4 个

数据统计量对  $\lg C_n^2$  进行分析,分别是平均绝对误差(MAE)、均方根误差(RMSE)、平均偏差(Bias)及相关系数( $R_{xy}$ )。4 个数据统计量具体表达式分别为

$$E_{MA} = \sum_{i=1}^N \frac{1}{N} |y_i - x_i|, \quad (15)$$

$$E_{RMS} = \sqrt{\sum_{i=1}^N \frac{1}{N} |y_i - x_i|^2}, \quad (16)$$

$$B_{bias} = \sum_{i=1}^N \frac{1}{N} (y_i - x_i), \quad (17)$$

$$R_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2 (y_i - \bar{Y})^2}}, \quad (18)$$

式中: $x_i$  代表某高度上实际测量值; $y_i$  代表同高度 SVM 模型的估算值; $N$  代表这一次探空 100 m 分层的数量; $\bar{X}$  为此次探空实际测量值的平均; $\bar{Y}$  为该次探空实验 SVM 模型估计的平均值。

最后误差统计结果如表 2 所示,可以看出:2018-05-05 的 SVM 模型估算的  $C_n^2$  廓线的平均绝对误差为 0.3722,均方根误差为 0.4461,平均偏差为 -0.2443,相关系数达到了 70.42%;2018-05-10 的 SVM 模型估算的  $C_n^2$  廓线的平均绝对误差为 0.3195,均方根误差为 0.3939,平均偏差为 -0.0476,相关系数达到了 62.17%。因此误差分析证明,根据以往少量测量数据训练出的 SVM 模型对以后  $C_n^2$  廓线的预测具有一定的可行性,具有一定泛化能力。

表 2 SVM 模型估算  $C_n^2$  廓线误差统计

Table 2 Error statistics of  $C_n^2$  profile estimated by SVM model

Data	MAE	RMSE	Bias	$R_{xy}$
2018-05-05	0.3722	0.4461	-0.2443	0.7042
2018-05-10	0.3195	0.3939	-0.0476	0.6217

图 8 和图 9 分别为 2018-05-05 和 2018-05-10 的  $C_n^2$  探空实测值和 SVM 模型估算值的频数分布直方图和累积概率分布图。图 8 和图 9 表明:在  $C_n^2$  探空廓线中, $\lg C_n^2$  的 SVM 模型估算值和探空实测值分布规律基本一致,符合正态分布特征。实测值和 SVM 模型估计值频率概率分布最大的区间为  $[-16, -15]$ ,当  $\lg C_n^2$  在  $[-16.5, -15.5]$  之间时, SVM 模型估算值的概率分布明显大于实测值,当  $\lg C_n^2 < -16.5$  或  $\lg C_n^2 > -15$  时, SVM 模型的估算值比实测值的概率分布小。

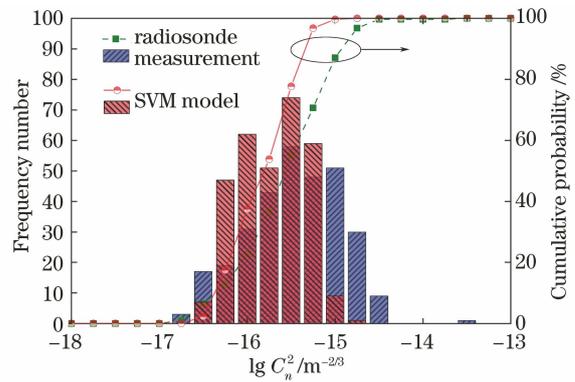


图 8 2018-05-05  $\lg C_n^2$  的频数分布直方图和累积概率密度分布图

Fig. 8 Frequency number distribution histogram and cumulative probability distribution of  $\lg C_n^2$  on 2018-05-05

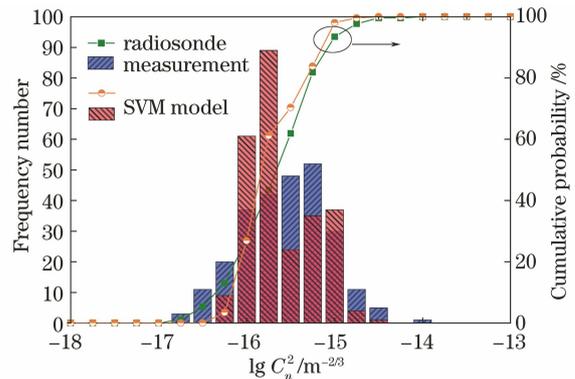


图 9 2018-05-10  $\lg C_n^2$  的频数分布直方图和累积概率密度分布图

Fig. 9 Frequency number distribution histogram and cumulative probability distribution of  $\lg C_n^2$  on 2018-05-10

图 10(a)是 2018-05-05 探空测量和 SVM 模型估算的  $C_n^2$  数值散点图,图 10(b)是 2018-05-10 探空测量和 SVM 模型估算的  $C_n^2$  数值散点图。经过研究,发现两天的估算结果都与探空测量值呈现出明显的线性关系,在 2018-05-05 出现略微向下偏移的现象,说明 SVM 模型对部分高层结的  $C_n^2$  值出现了低估,可能是因为  $C_n^2$  除了受常规气象参数影响外,还受间歇性和随机性等猝发机制的影响,这需要更多探空实验修正和验证 SVM 模型。而在 2018-05-10 的散点图中,数值较为均匀地分布在两侧,未出现低估情况。

## 4 结 论

建立 SVM 模型,使用常规气象参数廓线作为输入估算了沿海地区的  $C_n^2$  廓线,并将估算结果与

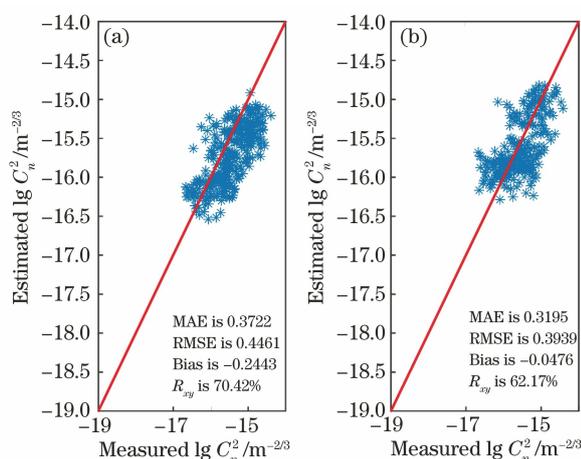


图 10 探空测量和 SVM 模型估算的  $\lg C_n^2$  数据散点图。

(a) 2018-05-05; (b) 2018-05-10

Fig. 10 Scatter diagrams of  $\lg C_n^2$  value obtained by radiosonde measurement and SVM model.

(a) 2018-05-05; (b) 2018-05-10

实测结果进行比较。SVM 估算模型基本能准确估算出  $C_n^2$  廓线的变化趋势,也揭示了光学湍流廓线的分层结构的基本特性——在自由大气中呈现指数递减和对流层增强的特征;定量来说,相关性分别达到了 70% 和 62%,均方根误差不超过 0.45,证明了 SVM 模型估算  $C_n^2$  廓线的可行性,且具有气象站常规探空数据直接估算  $C_n^2$  廓线的可能性。在估算  $C_n^2$  廓线时,与原始探空数据  $C_n^2$  廓线相比,SVM 模型出现了低估  $C_n^2$  值的现象。 $C_n^2$  值是多种因素同时作用的复杂结果,SVM 模型的低估有可能是忽略了对流、重力内波、湍流的间歇性和随机性猝发机制的影响。为了更精确地估算  $C_n^2$  廓线,需要加入更多能描述这些精细结构的物理量,这限制了 SVM 在该领域的优势,同时这些参数也是探空探测的难点所在,这也是大气光学工作者在 SVM 模型乃至探空探测领域未来可以进一步开展的工作。

### 参 考 文 献

[1] Rao R Z. Light propagation in the turbulent atmosphere [M]. Hefei: Anhui Science and Technology Press, 2005.  
饶瑞中. 光在湍流大气中的传播[M]. 合肥: 安徽科学技术出版社, 2005.

[2] Brown J H, Good R E. Thermosonde  $C_n^2$  measurement in Hawaii-Agust 1982 [R]. Massachusetts: Air Force Geophysics Laboratory, 1984.

[3] Good R E, Beland R R, Murphy E A, et al. Atmospheric models of optical turbulence [J].

Proceedings of SPIE, 1988, 0928: 165-186.

- [4] Hufnagel R E, Stanley N R. Modulation transfer function associated with image transmission through turbulent media[J]. Journal of the Optical Society of America, 1964, 54(1): 52-61.
- [5] van Zandt T E, Green J L, Gage K S, et al. Vertical profiles of refractivity turbulence structure constant: comparison of observations by the Sunset Radar with a new theoretical model[J]. Radio Science, 1978, 13(5): 819-829.
- [6] Tartaski B N. Wave propagation in a turbulent medium[M]. Beijing: Science Press, 1978.  
[苏]塔塔尔斯基. 湍流大气中波的传播理论[M]. 北京: 科学出版社, 1978.
- [7] Coulman C E, Vernin J, Coqueugniot Y, et al. Outer scale of turbulence appropriate to modeling refractive-index structure profiles [J]. Applied Optics, 1988, 27(1): 155-160.
- [8] Dewan E M, Good R E, Beland R, et al. A model for  $C_n^2$  (optical turbulence) profiles using radiosonde data [EB/OL]. [2021-05-07]. [https://www.researchgate.net/publication/235176586\\_A\\_Model\\_for\\_Csubn2\\_Optical\\_Turbulence\\_Profiles\\_Using\\_Radiosonde\\_Data](https://www.researchgate.net/publication/235176586_A_Model_for_Csubn2_Optical_Turbulence_Profiles_Using_Radiosonde_Data).
- [9] Han Y J, Wu X Q, Luo T, et al. Optical turbulence measurement over Lhasa using a meteorological radiosonde and model analysis [J]. Acta Optica Sinica, 2020, 40(6): 0601002.  
韩亚娟, 吴晓庆, 罗涛, 等. 拉萨光学湍流探空测量与模式分析[J]. 光学学报, 2020, 40(6): 0601002.
- [10] 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.
- [11] Lü J, Zhu W Y, Cai J, et al. Comparison of two approaches for estimating atmospheric optical turbulence intensity near sea[J]. Acta Optica Sinica, 2017, 37(5): 0501001.  
吕洁, 朱文越, 蔡俊, 等. 两种估算近海面大气光学湍流强度方法的比较[J]. 光学学报, 2017, 37(5): 0501001.
- [12] Chen X W, Zhu W Y, Qian X M, et al. Estimation of surface layer optical turbulence using artificial neural network [J]. Acta Optica Sinica, 2020, 40(24): 2401002.  
陈小威, 朱文越, 钱仙妹, 等. 基于人工神经网络的近地面光学湍流估算[J]. 光学学报, 2020, 40(24): 2401002.
- [13] Li H. Methods of statistical learning [M]. Beijing: Tsinghua University Press, 2012.  
李航. 统计学习方法[M]. 北京: 清华大学出版社,

- 2012.
- [14] Goodfellow I, Bengio Y, Courville A. Deep learning [M]. Zhao S J, Li Y J, Fu T F, et al., Transl. Beijing: Posts and Telecom Press, 2017.  
伊恩·古德费落, 加书亚·本吉奥, 亚伦·库维尔. 深度学习[M]. 赵申剑, 黎彧君, 符天凡, 等. 译. 北京: 人民邮电出版社, 2017.
- [15] Zhou Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016.  
周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [16] Xiao L M, Weng N Q, Ma C S, et al. Experimental research on the reliability of  $C_n^2$  sonde and its data processing method[J]. Journal of Hefei University of Technology (Natural Science), 2004, 27(2): 179-182.  
肖黎明, 翁宁泉, 马成胜, 等.  $C_n^2$  探空仪的可靠性实验与数据处理方法[J]. 合肥工业大学学报(自然科学版), 2004, 27(2): 179-182.
- [17] Kolmogorov A N. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers [J]. Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences, 1991, 434(1890): 9-13.