

基于梯度提升树的大气分层透过率快速计算方法^{*}

梁红丽 张 鹏 陈 林 马 刚 白文广 漆成莉

LIANG Hongli ZHANG Peng CHEN Lin MA Gang BAI Wenguang QI Chengli

1. 中国气象科学研究院, 北京, 100081

2. 国家卫星气象中心, 北京, 100081

1. *Chinese Academy of Meteorological Sciences, Beijing 100081, China*

2. *National Satellite Meteorological Center, Beijing 100081, China*

2019-08-30 收稿, 2020-04-30 改回.

梁红丽, 张鹏, 陈林, 马刚, 白文广, 漆成莉. 2020. 基于梯度提升树的大气分层透过率快速计算方法. 气象学报, 78(5): 853-863

Liang Hongli, Zhang Peng, Chen Lin, Ma Gang, Bai Wenguang, Qi Chengli. 2020. A Gradient Boosting Tree method for rapid calculation of level-to-space transmittances. *Acta Meteorologica Sinica*, 78(5):853-863

Abstract The calculation of atmospheric transmittance is the core of solving emitted infrared radiative transfer equations. As the layer transmittance of the atmosphere is parameterized by functions of the mean layer temperature, water vapor, ozone and other gas concentrations, RTTOV (Radiative Transfer for TOVS) is able to quickly and accurately compute level-to-space transmittance and top-of-atmosphere radiance. However, large computing errors have been found in some strong absorption bands, for example, the water vapor band. To solve this problem, a machine learning method called Gradient Boosting Tree (GBT) is utilized to compute transmittances in this paper. 1406 typical profiles have been selected from the ECMWF IFS-137 (European Centre for Medium-Range Weather Forecasts, the Integrated Forecast System, 137-level-profile) as training samples. The water vapor channel (7.33 μm) of the IRAS (InfraRed Atmospheric Sounder)/FY-3 (Fengyun 3 series) is selected for a case study. A fast model to compute clear-sky level-to-space transmittance and brightness temperature by the GBT method (GBT model hereafter) has been built. The calculated transmittances and brightness temperatures have been validated by the ground-truth. The comparison results show that the mean absolute errors (MAE) of clear-sky transmittances calculated by the GBT model are 0.0012 (land) and 0.0009 (ocean), while the mean squared logarithmic errors (MSLE) are 0.0215 and 0.0095 over land and ocean, respectively. These results are smaller than those calculated by RTTOV (MAEs of 0.0008 (land) and 0.0010 (ocean), MSLEs of 0.0135 (land) and 0.0227 (ocean)). In addition, the MAEs of clear-sky BT calculated by the GBT model are 0.0949 K (land) and 0.0634 K (ocean), and the root mean square errors (RMSE) are 0.1352 K (land) and 0.0831 K (ocean), respectively, which are smaller than those by RTTOV (MAEs of 0.1685 K (land) and 0.1466 K (ocean), RMSEs of 0.1794 K (land) and 0.1685 K (ocean)). The case study in this paper demonstrates that in the water vapor band of IRAS, the transmittance and brightness temperature predicted by the GBT are more accurate than those simulated by RTTOV. Machine learning method has the potential to improve the accuracy of the transmittance and brightness temperature in the water vapor band. It provides an optional solution to the fast radiation transfer calculation.

Key words Atmospheric transmittance, Infrared radiative transfer, Machine learning, Gradient Boosting Tree

摘 要 大气透过率的计算是红外辐射传输计算的核心, RTTOV(Radiative Transfer for TOVS)通过建立大气廓线中温度、水汽、臭氧和其他气体浓度等参数与卫星通道透过率的统计关系, 可实现卫星通道透过率和大气顶辐射率的快速准确计算。但在一些复杂吸收波段, 如水汽波段, RTTOV 的计算误差较大。为提高 RTTOV 在水汽敏感波段的计算精度, 利用机器学习中的梯度提

^{*} 资助课题: 国家重点研发计划项目(2018YFB0504900)、国家自然科学基金项目(41675036)。

作者简介: 梁红丽, 主要研究红外大气辐射传输。E-mail: HelenLiang1@outlook.com

通信作者: 张鹏, 主要从事气象卫星应用系统工程设计和卫星资料遥感应用工作。E-mail: zhangp@cma.gov.cn

升树(Gradient Boosting Tree, GBT)方法,选取从 ECMWF(European Centre for Medium-Range Weather Forecasts)的 IFS-137(The Integrated Forecast System, 137-level-profile)廓线集中挑选的 1406 条廓线和由此计算的透过率真值作为样本,选取风云三号气象卫星上搭载的红外分光计(InfraRed Atmospheric Sounder, IRAS)通道 12(7.33 μm)进行个例研究,分别建立陆地和海洋晴空大气等压面至大气层顶透过率的快速计算模型(GBT 模型)。通过和透过率、亮温真值的比较,验证了 GBT 模型。比较结果显示,GBT 模型预测的透过率平均绝对误差(Mean Absolute Error, MAE)为:陆地 0.0012,海洋 0.0009;均方对数误差(Mean Squared Logarithmic Error, MSLE)为:陆地 0.0215,海洋 0.0095,均小于 RTTOV 直接计算的透过率的误差(陆地、海洋的 MAE 分别比 RTTOV 小 0.0008 和 0.0010,MSLE 分别比 RTTOV 小 0.0135 和 0.0227);由 GBT 模型计算的亮温 MAE 分别为:陆地 0.0949 K,海洋 0.0634 K,均方根误差(Root Mean Square Error, RMSE)分别为:陆地 0.1352 K,海洋 0.0831 K,也都小于 RTTOV 直接模拟的晴空亮温误差(陆地、海洋的 MAE 分别比 RTTOV 小 0.1685 K 和 0.1466 K, RMSE 分别比 RTTOV 小 0.1794 K 和 0.1685 K)。本研究的结果表明,在 IRAS 红外水汽波段,GBT 预测的透过率和亮温误差比 RTTOV 小。机器学习有提高水汽波段正演精度的潜力,或可为辐射传输的快速计算提供可行的替代方法。

关键词 大气透过率, 红外辐射传输, 机器学习, 梯度提升树
中图法分类号 P407.6

1 引言

大气透过率的计算是红外辐射传输计算的核心(石广玉, 1998)。传统的大气透过率计算方法主要有 4 种:精确的逐线积分方法、近似的带模式方法(廖国男, 2004)、K 分布方法(石广玉, 1998; 张华, 1999),以及基于统计的 OPTRAN(Optical Path Transmittance)方案(McMillin, et al, 1976),根据这些方法,可将国际上存在的红外辐射传输模式分为精确模式和快速模式(张鹏等, 2005)两大类。

RTTOV(Radiative Transfer for TOVS)是一种快速辐射传输模式,它基于统计方案,由欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF)的 Eyre(1991)开发,最初用于变分系统。RTTOV 模拟大气透过率和大气顶亮温的基本原理是,基于典型大气廓线库的大气透过率逐线积分计算结果,建立卫星仪器通道透过率和大气状态、观测几何参数等因子的统计关系。它通过事先训练好的拟合系数,利用一组回归方程实现大气透过率的计算,计算效率高,且精度有保证,同时可以方便地用于云、气溶胶简化散射模型的快速处理(McMillin, et al, 2006)。随着卫星资料的大量增加,RTTOV 的需求也逐渐拓宽,如今已成为许多应用中必不可少的工具,如数值预报系统中的数据资料同化(Eyre, et al, 1993)、卫星数据的物理反演(Li, et al, 2000)、评估星上仪器的性能(Andrey-Andrés, et al, 2018)等。随着统计方案的改进和大气吸收谱线的升级,RTTOV 已多次更新,现已发布最新的 12.3 版。本研究选取 RTTOV

v12.1,用于计算等压面到大气顶的透过率和大气顶亮温。

具体地,RTTOV 利用实际大气廓线和卫星观测的几何参数、仪器通道的光谱特征参数等,建立光学厚度预报因子,通过统计回归,得到特定仪器和通道的预报因子系数,从而实现红外仪器通道透过率的快速精确计算(McMillin, et al, 1976)。但是,实际计算过程中,预报因子的选择依赖于经验,陆地的地表发射率等参数通常也假定为常数,这些固有设定使得 RTTOV 在复杂吸收波段(特别是水汽敏感波段)的计算精度有限。若减少对地表参数、光学厚度预报因子等的假设,可能会提高 RTTOV 在水汽通道的计算精度。

作为统计学习方法的延伸,机器学习对模型的可解释性要求降低,对模型结果的好坏关注更多,可在一定程度上自行选择数据特征,在实际任务中的表现通常也更好(Breiman, 2001)。常见的机器学习主要分监督学习、非监督学习、半监督学习和强化学习 4 类,监督学习主要有分类和回归两种任务。常见的回归方法有逻辑斯蒂回归、支持向量机、Bagging 和 Boosting 等。Boosting 是一种集成学习方法。集成学习是一种机器学习范式,和常见机器学习方法不同的是,集成学习通过学习一组“个体学习器”,再通过一定的策略将它们组合来完成学习任务。集成学习通常比单一的学习器泛化性能更好,对于“弱学习器”尤为明显(周志华, 2017)。

集成学习中的 Boosting 方法实际采用加法模型(基函数的线性组合)和前向分步算法实现学习的优化过程。以决策树为基函数的提升方法称为

提升树(Boosting Tree),它被认为是统计学习中性能最好的方法之一(李航, 2019)。对于损失函数为平方损失和指数损失的函数,分步优化的每一步都是简单的,但对一般的损失函数而言,每一步优化并不容易。针对这一问题,Friedman提出了梯度提升算法(Friedman, 2001, 2002)。梯度提升的关键是,利用损失函数的负梯度在当前模型的值作为回归问题提升树算法中残差的近似值(伪残差),拟合一个回归树。梯度提升树的快速、准确等优良特性使得它被广泛应用于各个领域,如光学特征识别、文本分类、人脸识别、基因表达分析等,任何机器学习派得上用场的领域,都可以尝试使用集成学习(周志华, 2017)。梯度提升树(Gradient Boosting Tree, GBT)以二叉树作为基学习器,通过它们的线性组合来拟合训练数据,即使输入和输出数据间的非线性关系很复杂,它的拟合效果也很好(李航, 2019)。近年来,很多利用机器学习和深度学习开展的应用研究表明,机器学习方法能够帮助解决很多气象和遥感问题,如强对流预报(郭瀚阳等, 2019)、冰雹预报(张文海等, 2019)和地面温度预报(嵇磊等, 2019)等。气象领域基于集成学习进行的研究也有不少,如对流初生预测(Liu, et al, 2019)、小麦病穗率预测(徐敏等, 2020)等。此外,Reid等(2015)用梯度提升树建立了野火预测模型,Yu等(2017)用空间加权的梯度提升树建立了PM_{2.5}的连续预测模型。在大气辐射领域,20世纪未曾涌现了一批神经网络在辐射传输领域的应用研究,主要包括大气廓线中温湿度、气体含量的卫星数据反演和大气顶辐射率及辐射通量的正演(Chevallier, et al, 1998, 2000; Schwander, et al, 2001)等。近年来,张华、白文广和张鹏等在大气辐射的快速计算研究方面取得了新的进展(白文广等, 2016, 2017; 张华, 2016; Zhang, et al, 2019),在红外辐射传输问题上,也有提高红外水汽通道正演精度的研究(曹玥瑶等, 2016)。但是,迄今为止,利用机器学习方法来提高红外水汽波段透过率计算精度的研究依然较少。红外水汽通道的吸收特性复杂,大气透过率的计算需要经过高度的非线性变换,梯度提升树方法作为高功能的学习算法,有望在红外水汽通道的透过率计算问题上发挥作用。因此,文中选取该方法,在RTTOV计算精度受限的水汽通道进行了研究,以期寻求一种更加准确且快

速的透过率计算方法。

2 研究方法

选取FY-3极轨系列气象卫星上搭载的红外分光计(IRAS, InfraRed Atmospheric Sounder)的通道12(7.33 μm),用GBT方法进行了大气分层透过率的学习。并将机器学习计算的透过率和RTTOV的结果进行比较,评估了GBT方法的表现。

2.1 梯度提升树

假定输入数据为 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_i \in X \subseteq R^n$, $y_i \in Y \subseteq R$, 对于回归问题, 如果将输入空间 X 划分为 J 个互不相交的区域 $R_1, R_2, \dots, R_J$, 且在每个区域上确定输出的常量为 c_j , 那么一棵回归树可表示为

$$T(x; \theta) = \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (1)$$

式中, 参数 $\theta = \{(R_1, c_1), (R_2, c_2), \dots, (R_J, c_J)\}$ 表示树的区域划分(R_j)和各区域上的常数, J 为回归树的复杂度,即叶结点数。

对于回归问题, 梯度提升的算法(李航, 2019)如下:

输入: 训练数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_i \in X \subseteq R^n$, $y_i \in Y \subseteq R$, 损失函数 $L(y, f(x))$;

输出: 回归树 $F(x)$

(1)初始化

$$F_0(x) = \arg \min_c \sum_{i=1}^n L(y_i, c) \quad (2)$$

(2)对 $m = 1, 2, \dots, M$

a. 对 $i = 1, 2, \dots, n$, 计算伪残差

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad i = 1, 2, \dots, n \quad (3)$$

b. 对 r_{im} 拟合一个回归树, 得到第 m 棵树的叶结点区域 R_{mj} , $j = 1, 2, \dots, J$

c. 对 $j = 1, 2, \dots, J$, 计算

$$c_{mj} = \arg \min_c \sum_{x_i \in R_{mj}} L(y_i, F_{m-1}(x_i) + c) \quad (4)$$

d. 更新

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (5)$$

(3)得到回归树

$$F(x) = F_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (6)$$

采用平方误差损失函数时,拟合的伪残差即为真实残差, $r_{im} = y_i - F_{m-1}(x_i)$ 。

式中, $L(y_i, c)$ 为样本 x_i 的损失, 式(2)表示在树生长初始的划分, 选择使训练集样本损失函数最小的划分 c , 此时回归树为只有一个根结点的树; 式(3)即为树的梯度计算, 作为伪残差, 用于一棵新树的生长(对 r_{im} 拟合回归树), 式(4)计算新树的划分常量 c , 估计叶结点区域, 拟合伪残差; 式(5)为树的更新, 式(6)为最终得到的回归树。回归树的学习流程如图1所示, 从图中可以直观地看出学习每一棵树的过程。

针对大气透过率的计算, GBT 的输入为大气廓线各等压面的特征量, 输出为各等压面到大气顶的透过率。训练数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_i \in X \subseteq R^{n \times m}$, $y_i \in Y \subseteq R^l$, $n=1406$, 为廓线条数, $m=332$, 为每条廓线的特征数, 包括各等压面的温度、水汽含量、臭氧浓度和近地面 2 m 高处的气温、地

表温度, 以及参考 RTTOV 的光学厚度预报因子计算的 27 个因子, 这些因子的计算方法具体可见 (Matricardi, 2008), $l=101$, 为每条大气廓线垂直分层的数目。

GBT 的训练用 Python 的机器学习扩展包 sklearn 实现, 损失函数采用平方根损失, 选取最佳参数。重点调整的参数有集成的决策树的数目 ($n_estimators$)、每棵树的深度 (max_depth) 和学习率, 即梯度下降的步长 ($learning_rate$)。

2.2 评估标准

RTTOV 在透过率计算上的表现常通过用每个等压面上和逐线积分模式 LBLRTM (Line-by-line Radiative Transfer Model) 的偏差 ($bias_j$) 及均方根误差 (Root Mean Squared Error, $RMSE_j$) 进行评估 (Saunders, et al, 1999; Singh, et al, 2009), 文中借鉴此方法, 对 GBT 模型的表现做评估。此外, GBT 在所有等压面的整体表现以平均绝对误差 (Mean Absolute error, MAE) 和均方对数误差 (Mean Squared Logarithmic Error, MSLE) 为评估标准。均方对数误差的定义如下

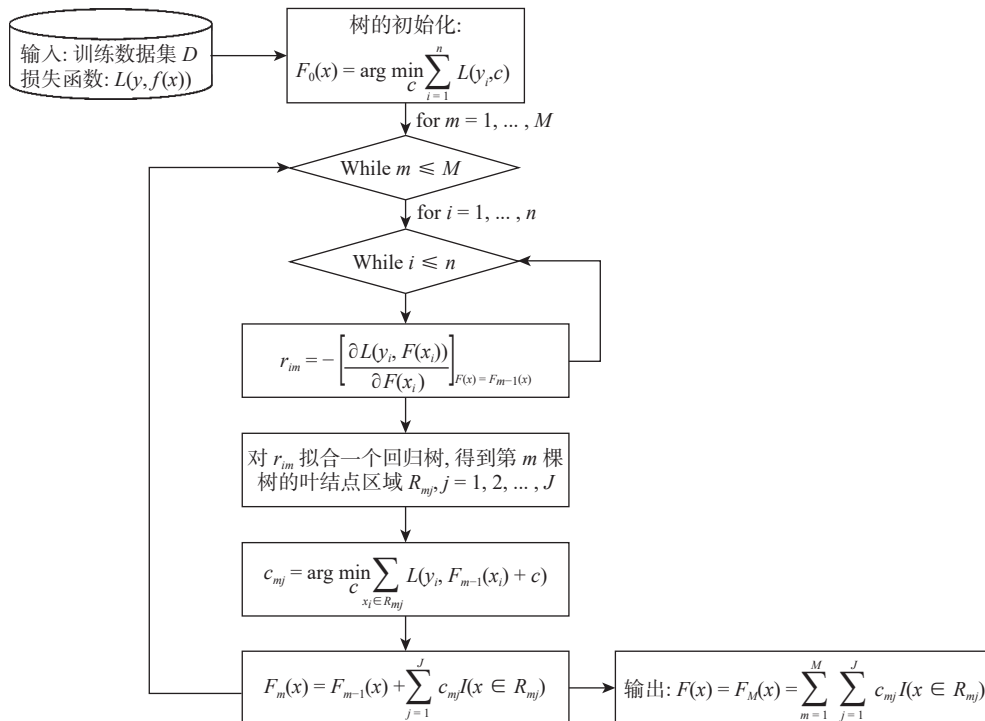


图1 GBT 的学习流程

Fig. 1 Working flow chart of GBT

$$\text{MSLE} = 10^4 \times \frac{1}{n} \frac{1}{N} \sum_{j=1}^n \sum_{i=1}^N (\ln(1 + y_{ij}) - \ln(1 + \hat{y}_{ij}))^2 \quad (7)$$

由于各等压面到大气顶的透过率呈非线性变化,且接近指数形分布(图2),因此,计算一条廓线在所有等压面的透过率误差时,平均绝对误差可能受极端值影响较大,而均方对数误差可在一定程度上减小非线性变化对评估结果的影响。由于透过率本身在[0, 1]范围内,取值较小,因此直接计算的均方对数误差很小。式(7)中,10⁴和标准MSLE计算公式相乘是为了方便查看结果。

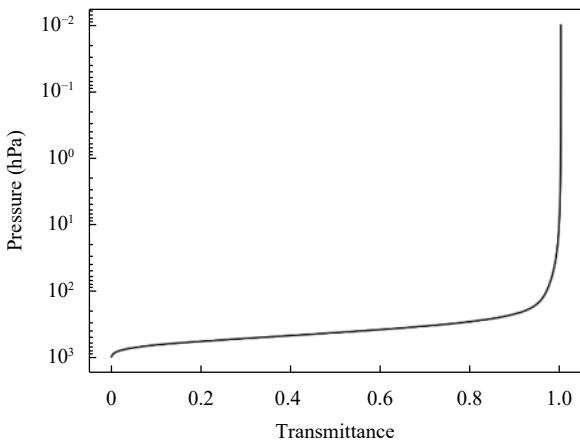


图2 大气分层透过率的垂直变化
Fig. 2 Vertical variation of atmospheric level-to-space transmittance

这4个评估标准中, bias_j 和 RMSE_j 可查看GBT在第 j 个等压面上计算透过率的准确性,平均绝对误差、均方对数误差可以综合评估GBT在数据集的所有廓线、所有等压面上的综合表现。

3 遥感仪器和数据

3.1 遥感仪器

IRAS是FY-3极轨系列卫星上搭载的红外探测器,在0.69—15.0 μm范围设计有26个光谱通道,包括20个红外通道、6个可见光和近红外通道。IRAS是用于观测地球天气和环境的重要仪器,观测数据主要用于数值天气预报中的卫星资料同化和全球大气的温湿廓线、臭氧含量和云参数的反演等(Qi, et al, 2012),其工作原理和搭载于Metop上的红外辐射探测器HIRS(High-resolution Infrared Radiation Sounder)相似。

Garand等(2001)和Roger Saunders等(2018)分别对RTTOV在HIRS各通道的表现做过评估。结果表明,RTTOV在非水汽通道的正演误差很小,在水汽敏感通道的正演误差较大。Saunders等(2018)用第7和8版的预报因子计算的结果表明,RTTOV和LBLRTM的差异都在11(7.326 μm)、12(6.523 μm)通道最大,达0.1 K。IRAS上同样设置了7.33和6.52 μm两个水汽敏感通道,RTTOV在两个水汽敏感通道的计算误差明显大于其他通道。图3为通道12的光谱响应函数曲线,图4为不同波段的大气分子吸收强度,结合两幅图可以看出,在7.33 μm(约1364 cm⁻¹)处,除水汽的强吸收外,还有N₂O、CH₄等分子的吸收,增加了准确计算该通道透过率的难度。因此,选取IRAS通道12

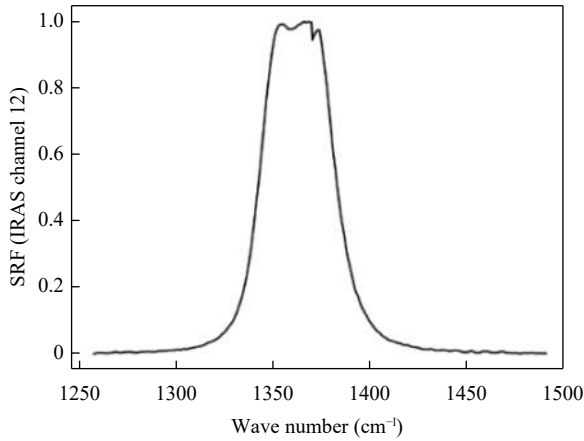


图3 IRAS通道12的光谱响应函数
Fig. 3 Spectral response function of IRAS channel 12

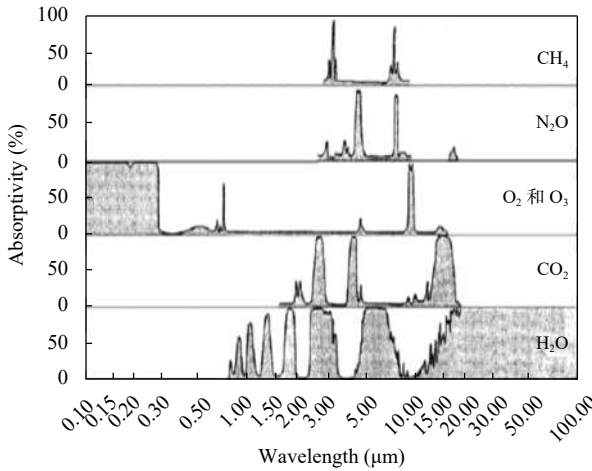


图4 大气的分子吸收(Peixoto, et al, 1992)
Fig. 4 Atmospheric molecular absorption
(Peixoto, et al, 1992)

(7.33 μm)作为GBT的学习通道,以期提高RTTOV在该通道上透过率计算的准确率。

获取大气分层透过率的真值时,选取高光谱分辨率红外大气探测仪IASI(Infrared Atmospheric Sounding Interferometer)。透过率真值的具体计算过程在3.3节做详细介绍。

3.2 大气廓线数据

大气廓线集是透过率计算的基础数据,要求具有代表性,一般从无线电探空或数值预报分析场中遴选获得,文中使用的大气廓线从IFS-137(Integrated Forecasting System, 137-level profile)集合中选取。

ECMWF先后建立过将大气垂直分为31、50和60个等压面的廓线集。Chevallier等(2006)提出新的廓线取样策略,在阈值挑选的基础上,对不同变量进行了独立取样,从ECMWF预报系统的cycle 30R2生成的廓线中进行挑选,将廓线更新升级到了91个等压面,并建立了按照温度、相对湿度、臭氧浓度、云水含量、降水5个变量取样做成的5个廓线集IFS-91。之后,2014年,ECMWF以相似的策略,增大随机选取的比例后,用2013年9月至2014年8月的预报场制作了最新的IFS-137廓线集(Eresmaa, et al, 2014)。与IFS-91相比,IFS-137廓线集在时间和空间分布上更加均匀,各变量的垂直分布也更接近当前的状态。因此,文中基于该廓线集,选取其中南北纬30°内的2588条晴空廓线。这2588条廓线中的垂直廓线信息包括气压、温度、水汽含量、臭氧含量和近地面温度等,能够满足RTTOV对输入廓线的信息需要。

3.3 数据处理

RTTOV最新制作的系数文件将大气分为101个等压面,大气顶的气压值为0.005 hPa,地表气压设置为1100 hPa,各等压面的具体气压值见Matricardi

(2008)。文中所选廓线的大气分层数是137,因此,首先将廓线各变量插值到101个等压面。插值时,如果地面气压小于1100 hPa,则在高于地面气压的等压面上对温度和臭氧廓线做常数外插,水汽廓线做正常的插值处理。由于RTTOV对所有廓线输入量规定了常数阈值,插值得到的集合中有超出阈值的廓线,因此,根据RTTOV中输入廓线各变量的阈值,再次进行筛选,得到2576条廓线,以各等压面变量的0.3和99.7为分位点,剔除超出分位数的廓线,最终得到1406条,其中陆地廓线443条,海洋廓线963条。这些廓线的空间分布情况见图5。图6a、c和e分别给出筛选前原始廓线的温度、水汽和臭氧廓线在每个等压面的均值、上下四分位数、上下十分位数及上下0.3分位数的分布。分析这些廓线可以发现,无论温度还是臭氧和水汽廓线,几乎每个等压面都有和整体分布差异较大的廓线,即两边的极大值和极小值和上下0.3分位数仍有相当的差距。图6b、d和f给出的是筛选后廓线的温度、水汽、臭氧垂直分布。可以看出,最终选出的1406条廓线在垂直分布上都更加均匀合理,在空间分布上也较均匀。

廓线集制作完成后,计算大气各廓线的大气分层透过率真值。由于LBLRTM计算成百上千条廓线的透过率需要大量时间和存储空间,IASI的观测精确度高,光谱连续,且光谱分辨率高于IRAS通道12的光谱响应函数,因此文中选取了一种替代方法:用RTTOV计算所选廓线集在晴空条件下IASI的等压面到大气顶的大气透过率作为单色透过率,利用IRAS的通道光谱响应函数,将单色透过率卷积至IRAS通道12,计算通道透过率,作为GBT所学透过率的真值。用RTTOV计算晴空条件下IRAS通道12的等压面到大气顶的透过率和真值比较,其误差即为GBT需要达到并减小的目标误差。

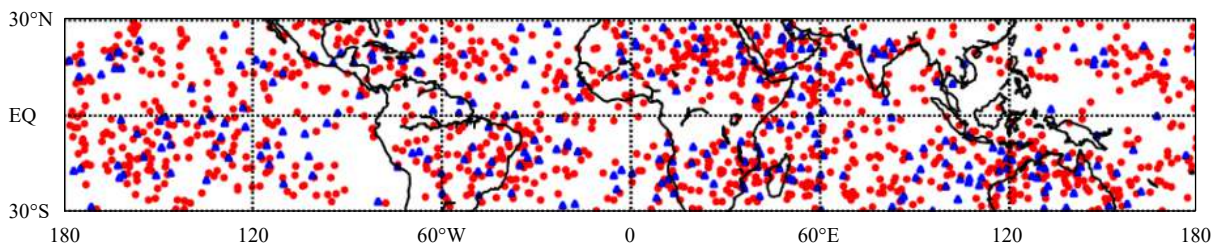


图5 1406条廓线的分布(红色点为训练集廓线,蓝色三角为测试集廓线)

Fig. 5 Distribution of the selected 1406 profiles (red dots: profiles of training set, blue triangles: profiles of testing set)

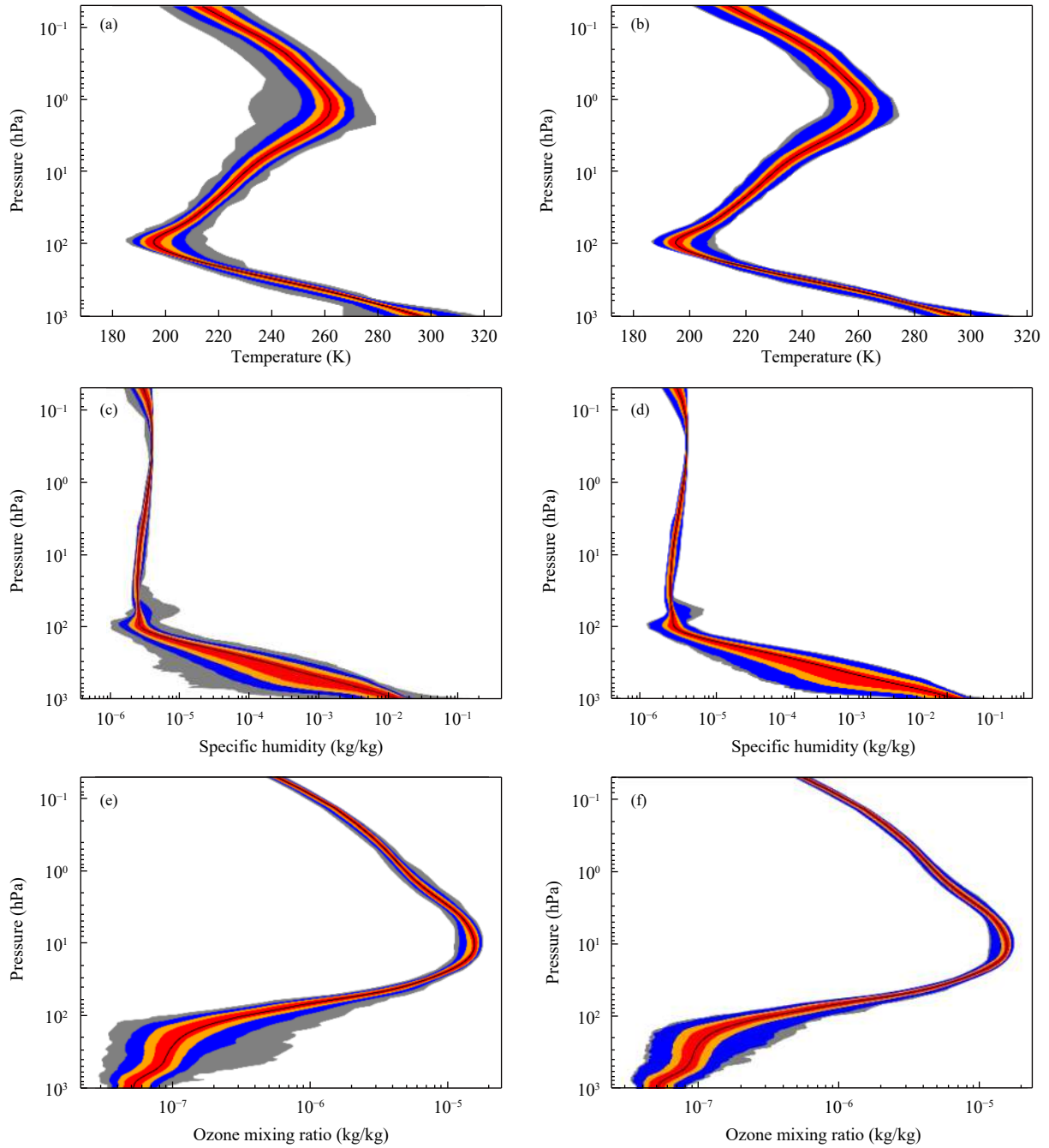


图6 原始廓线 (a、c、e) 和筛选得到的 1406 条廓线 (b、d、f) 的温度 (a、b)、湿度 (c、d)、臭氧 (e、f) 垂直分布 (灰色为极大、极小值包含的区域, 蓝色为上下 0.3 分位数包含的区域, 橙色为上下十分位数包含的区域, 红色为上、下四分位数包含的区域, 黑色线为均值廓线)

Fig. 6 Distributions of temperature (a, b), specific humidity (c, d) and ozone (e, f) in the original (a, c, e) and 1406 profiles database (b, d, f) (Gray shading indicates the range constrained by minimum and maximum values, blue shading are for values constrained by 0.3th and 99.7th percentiles, orange shadings indicate values constrained by 10th and 90th percentiles, and red shading indicates values constrained by 25th and 75th percentiles. Black solid line shows the mean profile)

由 RTTOV 计算的 IASI 透过率作为光谱分辨率为 0.25 cm^{-1} 的单色透过率, 卷积至 IRAS 通道 12 作为透过率的真值, 辐射率作为单色辐射率, 卷

积至 IRAS 通道 12 后(辐射率的卷积公式如式(8)), 用普朗克反函数计算亮温作为亮温的真值; IRAS 的透过率和亮温均作为参考用于比较。

$$R_v = \frac{\int R(v, \theta) \phi_v dv}{\int \phi_v dv} \quad (8)$$

式中, $R(v, \theta)$ 为单色透过率或辐亮度, ϕ_v 为归一化的通道光谱响应函数。

4 计算结果及分析

4.1 透过率的计算和比较

1406 条廓线集中包含了 443 条陆地廓线和 963 条海洋廓线, 由于陆地下垫面较复杂, 陆地廓线和海洋廓线有一定差异, 因此, 本研究分别训练了海洋和陆地的透过率计算模型, 均取 80% 样本用于训练, 20% 用于测试。对于海洋, 当决策树的数目为 180, 树的深度为 3, 学习率为 0.045 时, 得到的 4 种误差都最小; 对于陆地, 当决策树的数目为

150, 树的深度为 3, 学习率为 0.045 时, 得到的模型表现最好。海洋、陆地两种模型在测试集的透过率计算误差如图 7 和图 8 所示。图 7a、b 分别是 GBT 对陆地和海洋预测的透过率、RTTOV 模拟的透过率和真值的对比, 图 8a、b 分别是 GBT 和 RTTOV 在海洋和陆地预测的透过率在各等压面的误差对比。

图 7 的散点表明, 无论是陆地还是海洋, 从所有等压面的透过率来看, GBT 预测的透过率和真实值都相当接近。陆地的平均绝对误差仅 0.0012, 均方根误差为 0.0034, 均方对数误差为 0.0215, 均小于 RTTOV 的计算误差。从图 8 可以看出, 测试集上, 陆地 GBT 模型预测的大气分层透过率在绝大多数等压面的平均偏差都小于 0.0025, 均方根误差大多小于 0.01; 海洋 GBT 模型预测的大气分层透

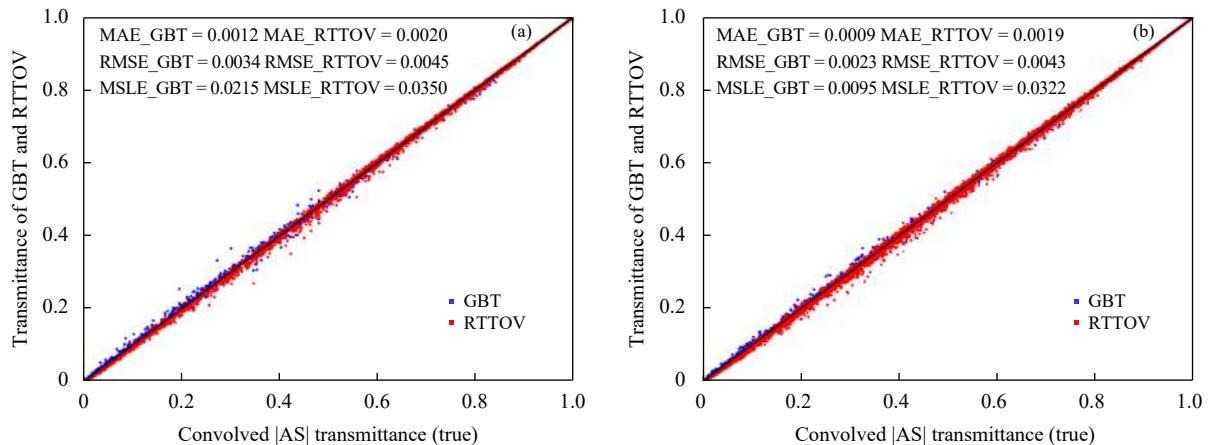


图 7 真实透过率和预测值对比 (a. 陆地, b. 海洋)

Fig. 7 GBT predicted transmittances and ground-truth transmittances (a. land, b. ocean)

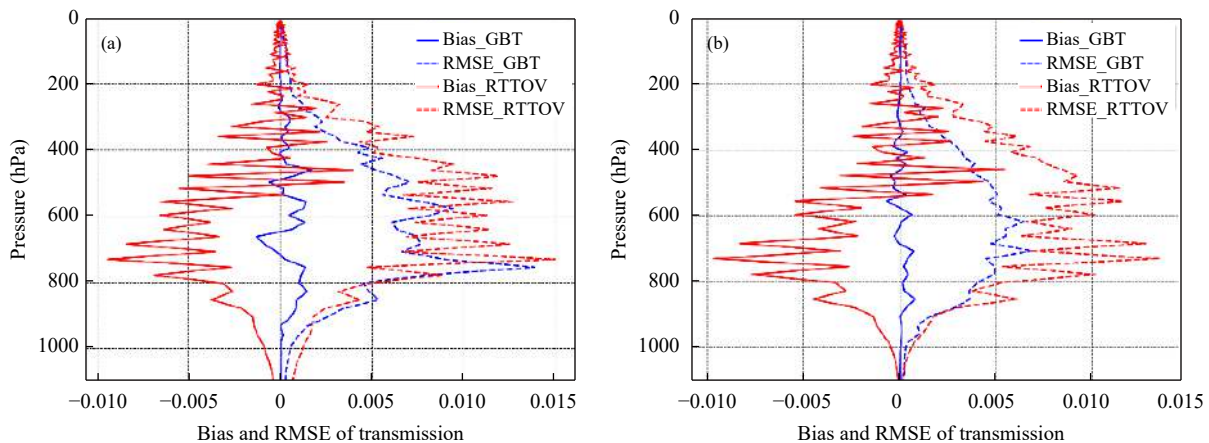


图 8 各等压面透过率的偏差和均方根误差 (a. 陆地, b. 海洋)

Fig. 8 Bias and RMSE of transmittance on different pressure levels (a. land, b. ocean)

过率在每个等压面的平均偏差都小于 0.0025, 均方根误差大多接近或小于 0.005。与 RTTOV 相比, 陆地和海洋的 GBT 模型预测的透过率在绝大多数等压面上的误差都更小, 其中, 海洋模型的表现比陆地模型更好, 这可能由陆地复杂的下垫面造成。其中, 在 700 hPa 上下的高度, 误差相对较大, 这和 7.33 μm 通道的特点一致。

4.2 亮温的计算和比较

用 IRAS 通道 12 的光谱响应函数对 IASI 的单个辐亮度进行卷积计算后, 通过普朗克逆函数计算出亮温作为亮温比较的真值, 记作 BT_true。基于 RTTOV 计算红外通道晴空亮温的方法, 利用 GBT 预测的透过率计算亮温, 记作 BT_pred。将 RTTOV 利用 IRAS 的预报因子系数计算出的亮温记作 BT_rttov。以 BT_true 为基准, 计算 BT_pred 和 BT_rttov

的偏差。图 9 为测试集中陆地模型预测的亮温偏差和海洋模型预测的亮温偏差。

从图中可以看出, 对于陆地和海洋, BT_rttov 和 BT_true 间的亮温偏差都在 -0.2 K 左右波动, BT_pred 和 BT_true 间的偏差在 0 K 左右波动, 说明由 GBT 预测的透过率计算出的亮温整体偏差更小。另外, 也可以看出, BT_pred 和 BT_true 间的偏差波动幅度更小, 特别是海洋模型, 上下波动的范围不超过 0.4 K, 陆地的波动幅度稍大, 但比 BT_rttov 和 BT_true 间的亮温偏差波动小。BT_rttov 在陆地模型、海洋模型测试集的平均绝对误差分别为 0.2634 和 0.2100 K, 均方根误差分别为 0.3146 和 0.2516 K, 而 BT_pred 在测试集的平均绝对误差为 0.0949 和 0.0634 K, 均方根误差为 0.1352 和 0.0831 K, 平均绝对误差分别比 BT_rttov 的小 0.1685 和 0.1466 K。

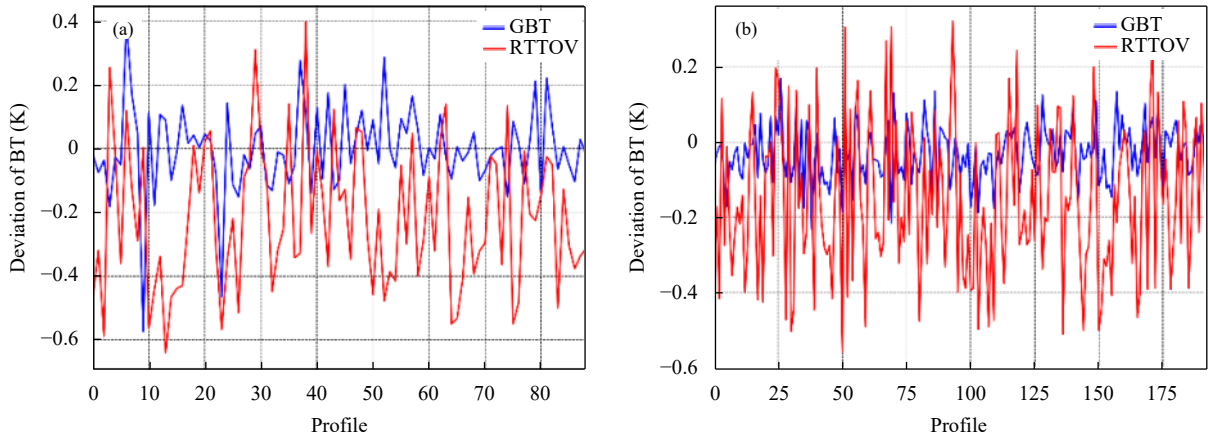


图 9 陆地模型(a)与海洋模型(b)GBT 和 RTTOV 在测试集的亮温偏差

Fig. 9 Deviation of brightness temperature computed by GBT and RTTOV with training profiles (a. land, b. ocean)

表 1 为 GBT 和 RTTOV 在透过率和亮温上的误差及平均用时, 总体来看, 无论透过率的计算还是亮温的计算, GBT 计算的误差都小于 RTTOV, 只

是在每条廓线上的计算时间略有增加。这种程度的时间增加在目前的计算机并行技术下将不会带来很大的问题。

表 1 GBT 和 RTTOV 的透过率和亮温误差比较

Table 1 MAE and MSLE of transmittance, MAE and RMSE of BT on testing set

方法	MAE- τ_{ao}		MSLE- τ_{ao}		MAE-BT(K)		RMSE-BT(K)		时间 (s)
	陆地	海洋	陆地	海洋	陆地	海洋	陆地	海洋	
RTTOV	0.0020	0.0019	0.0350	0.0322	0.2634	0.2100	0.3146	0.2516	0.170
GBT	0.0012	0.0009	0.0215	0.0095	0.0949	0.0634	0.1352	0.0831	0.849

5 结论和讨论

针对 RTTOV 在水汽敏感通道的透过率计算误差较大的问题, 利用机器学习中的梯度提升树方

法, 以 IRAS 的通道 12 为例, 选取 1406 条廓线, 试验了 GBT 在计算大气分层透过率上的效果。结果表明, 与 RTTOV 相比, GBT 计算大气分层透过率的准确率明显提高, 平均偏差远小于由 RTTOV 直

接计算的透过率。在亮温上的比较结果也表明,由 GBT 计算的透过率所得到的亮温误差更小。这对提高 RTTOV 正演模拟的精度有很大作用,也表明机器学习方法在复杂的大气辐射传输计算中可有用武之地。需要注意的是,GBT 对样本较敏感,通过适当的采样方法(如增大样本量、增大采样的随机性)实现数据样本扰动、增大样本的代表性,可以使 GBT 透过率计算模型适用性更强。

在计算透过率和亮温时,RTTOV 的误差主要来自两个方面:(1)建立系数时的统计误差,(2)插值误差。基于统计的 RTTOV 在每个等压层都使用相同的预先制定好的预报因子来进行计算光学厚度,再计算透过率,且对于红外波段,陆地地表的发射率都假设为常量。预报因子、插值、固定的地表发射率等设定都是 RTTOV 进行正演模拟的误差来源,对于水汽敏感通道的计算,误差更大。而机器学习方法并不需要做很多假定,因此可能获得更好的计算结果。

不过,文中所做试验仍有进一步优化和提升的空间。首先,GBT 学习的目标值是由 RTTOV 计算 IASI 的透过率后作为单色透过率、卷积至 IRAS 通道得到,与 LBLRTM 相比精度会有所降低。但是鉴于 LBLRTM 计算成百上千条廓线的透过率计算量非常大、耗时长,文中主要探讨 GBT 在红外水汽波段计算透过率的可行性,因此选取了这样的替代方法。如果能有多条廓线的 LBLRTM 结果做目标值,学习的结果将更为可靠,工作的进一步优化可选择使用 LBLRTM。其次,目前只选取了 IRAS 的 $7.33\text{ }\mu\text{m}$ 通道来试验 GBT 改进 RTTOV 水汽通道模拟精度的可行性,廓线也只选取了 30°N — 30°S 的大气廓线,这些廓线的水汽含量相对较高,且水汽和温度垂直变化相对较缓和,利于对水汽通道做训练和检验,要扩大正演模拟的范围,可能还需要增加其他纬度带的廓线。未来将在包括水汽和非水汽等多个通道进行模拟,并扩大廓线的分布范围,进一步验证 GBT 相比 RTTOV 的优势。如果根据通道的权重函数相应地选取地表、光谱等信息,作为 GBT 的输入特征,可能会获得更好的结果。无论如何,文中研究结果已经展示了机器学习在辐射传输计算方面的优势和应用潜力。

文中初步验证了 GBT 在红外水汽通道的正演模拟优于 RTTOV,若在其他通道的正演精度和

RTTOV 也不相上下,可在长远工作上考虑将 GBT 计算透过率的模型集成,替换传统辐射传输模式的上行辐射计算模块。

致谢:感谢哈尔滨工业大学(深圳)计算机学院叶允明老师、李旭涛老师和他们的团队,在研究过程中,他们热心给予了方法上的重要指导和帮助。

参考文献

- 白文广,张鹏,张文建等. 2016. 一种高效计算高光谱分辨率红外大气辐射传输的方法. *红外与毫米波学报*, 35(1): 99-108. Bai W G, Zhang P, Zhang W J, et al. 2016. An efficient method for hyper-spectral infrared atmospheric radiation transfer calculation. *J Infrared Millim Waves*, 35(1): 99-108 (in Chinese)
- 白文广,张鹏,张文建等. 2017. 遥感用途的高光谱中短波红外辐射传输精确计算. *中国科学: 地球科学*, 47(12): 1483-1492. Bai W G, Zhang P, Zhang W J, et al. 2017. A model for accurately calculating hyper-spectral, middle-shortwave infrared radiative transfer for remote sensing. *Sci China Earth Sci*, 47(12): 1483-1492 (in Chinese)
- 曹玥瑶,张鹏,马刚等. 2016. FY-3 IRAS 水汽通道亮温正演精度改进方法. *应用气象学报*, 27(6): 698-708. Cao Y Y, Zhang P, Ma G, et al. 2016. An improvement of brightness temperature simulation of FY-3 IRAS infrared water vapor channel. *J Appl Meteor Sci*, 27(6): 698-708 (in Chinese)
- 郭瀚阳,陈明轩,韩雷等. 2019. 基于深度学习的强对流高分辨率临近预报试验. *气象学报*, 77(4): 715-727. Guo H Y, Chen M X, Han L, et al. 2019. High resolution nowcasting experiment of severe convections based on deep learning. *Acta Meteor Sinica*, 77(4): 715-727 (in Chinese)
- 嵇磊,王在文,陈敏等. 2019. 人工智能技术能否提高地面气温预报的精度: 记 AI Challenger 2018 全球天气预报挑战赛. *气象学报*, 77(5): 960-964. Ji L, Wang Z W, Chen M, et al. 2019. How much can AI techniques improve surface air temperature forecast: A report from AI Challenger 2018 Global Weather Forecast Contest. *Acta Meteor Sinica*, 77(5): 960-964 (in Chinese)
- 李航. 2019. 统计学习方法. 2 版. 北京: 清华大学出版社, 166-171. Li H. 2019. *Statistical Learning Method*. 2nd ed. Beijing: Tsinghua University Press, 166-171 (in Chinese)
- 廖国男,郭彩丽,周诗健,译. 2004. 大气辐射导论. 北京: 气象出版社. Liao K N, Guo C L, Zhou S J, trans. 2004. *Introduction to Atmospheric Radiation*. Beijing: China Meteorological Press (in Chinese)
- 石广玉. 1998. 大气辐射计算的吸收系数分布模式. *大气科学*, 22(4): 659-676. Shi G Y. 1998. On the k-distribution and correlated-distribution models in the atmospheric radiation calculations. *Chinese J Atmos Sci*, 22(4): 659-676 (in Chinese)
- 徐敏,徐经纬,谢志清等. 2020. 随机森林机器学习算法在江苏省小麦赤霉病病穗率预测中的应用. *气象学报*, 78(1): 143-153. Xu M, Xu J W, Xie Z Q, et al. 2020. Application of the random forest machine algorithm in

- forecasting diseased panicle rate of wheat scab in Jiangsu province. *Acta Meteor Sinica*, 78(1): 143-153 (in Chinese)
- 张华. 1999. 非均匀路径相关 K-分布方法的研究 [D]. 北京: 中国科学院大气物理研究所. Zhang H. 1999. On the study of a new correlated K-distribution method for nongray gaseous absorption in the inhomogeneous scattering atmosphere[D]. Beijing: The Institute of Atmospheric Physics, Chinese Academy of Sciences (in Chinese)
- 张华. 2016. BCC_RAD 大气辐射传输模式. 北京: 气象出版社. Zhang H. 2016. BCC_RAD Radiative Transfer Model. Beijing, China Meteorological Press (in Chinese)
- 张鹏, Brunel P, 董超华等. 2005. 卫星高光谱红外大气探测的正演模拟研究. *红外与毫米波学报*, 24(6): 414-418. Zhang P, Brunel P, Dong C H, et al. 2005. Study of hyperspectral IR atmospheric sounding with an accurate forward simulation. *J Infrared Millim Waves*, 24(6): 414-418 (in Chinese)
- 张文海, 李磊. 2019. 人工智能在冰雹识别及临近预报中的初步应用. *气象学报*, 77(2): 282-291. Zhang W H, Li L. 2019. A preliminary application of artificial intelligence on the detection and nowcasting of hail weather. *Acta Meteor Sinica*, 77(2): 282-291 (in Chinese)
- 周志华. 2017. 机器学习. 北京: 清华大学出版社, 171-178. Zhou Z H. 2017. Machine Learning. Beijing: Tsinghua University Press, 171-178 (in Chinese)
- Andrey-Andrés J, Fourrié N, Guidard V, et al. 2018. A simulated observation database to assess the impact of the IASI-NG hyperspectral infrared sounder. *Atmos Meas Tech*, 11(2): 803-818
- Breiman L. 2001. Statistical modelling: The two cultures. *Stat Sci*, 16(3): 199-231
- Chevallier F, Chérut F, Scott N A, et al. 1998. A neural network approach for a fast and accurate computation of a longwave radiative budget. *J Appl Meteor*, 37(11): 1385-1397
- Chevallier F, Morcrette J J, Chérut F, et al. 2000. Use of a neural-network-based long-wave radiative-transfer scheme in the ECMWF atmospheric model. *Quart J Roy Meteor Soc*, 126(563): 761-776
- Chevallier F, Di Michele S, McNally A. 2006. Diverse profile datasets from the ECMWF 91-level short-range forecasts. NWP SAF Report No. NWPSAF-EC-TR-010, Shinfield Park, Reading: ECMWF
- Eresmaa R, McNally A P. 2014. Diverse profile datasets from the ECMWF 137-level short-range forecasts. Document, NWPSAF_EC_TR_017. Shinfield Park, Reading: ECMWF
- Eyre J R. 1991. A fast radiative transfer model for satellite sounding systems. ECMWF, Technical Memorandum. 176. Shinfield Park, Reading: ECMWF
- Eyre J R, Kelly G A, McNally A P, et al. 1993. Assimilation of tovs radiance information through one-dimensional variational analysis. *Quart J Roy Meteor Soc*, 119(514): 1427-1463
- Friedman J H. 2001. Greedy function approximation: A gradient boosting machine. *Ann Stat*, 29(5): 1189-1232
- Friedman J H. 2002. Stochastic gradient boosting. *Comput Stat Data Anal*, 38(4): 367-378
- Garand L, Turner D S, Larocque M, et al. 2001. Radiance and Jacobian intercomparison of radiative transfer models applied to HIRS and AMSU channels. *J Geophys Res: Atmos*, 106(D20): 24017-24031
- Li J, Wolf W W, Menzel W P, et al. 2000. Global soundings of the atmosphere from atovs measurements: The algorithm and validation. *J Appl Meteor*, 39(8): 1248-1268
- Liu Z J, Min M, Li J, et al. 2019. Local severe storm tracking and warning in pre-convection stage from the new generation geostationary weather satellite measurements. *Remote Sens*, 11(4): 383
- Matricardi M. 2008. The generation of RTTOV regression coefficients for IASI and AIRS using a new profile training set and a new line-by-line database. Technical Memorandum, 564, EUMETSAT Contract No. EUM/CO/06/1504/PS. Shinfield Park, Reading: ECMWF
- McMillin L M, Fleming H E. 1976. Atmospheric transmittance of an absorbing gas: A computationally fast and accurate transmittance model for absorbing gases with constant mixing ratios in inhomogeneous atmospheres. *Appl Opt*, 15(2): 358-363
- McMillin L M, Xiong X Z, Han Y, et al. 2006. Atmospheric transmittance of an absorbing gas. 7: Further improvements to the OPTRAN 6 approach. *Appl Opt*, 45(9): 2028-2034
- Peixoto J P, Oort A H. 1992. Physics of Climate. New York, USA: American Institute of Physics
- Qi C L, Chen Y, Liu H, et al. 2012. Calibration and validation of the Infrared atmospheric sounder onboard the FY-3B satellite. *IEEE Trans Geosci Remote Sens*, 50(12): 4903-4914
- Reid C E, Jerrett M, Petersen M L, et al. 2015. Spatiotemporal prediction of fine particulate matter during the 2008 Northern California wildfires using machine learning. *Environ Sci Technol*, 49(6): 3887-3896
- Saunders R, Matricardi M, Brunel P. 1999. An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quart J Roy Meteor Soc*, 125(556): 1407-1425
- Saunders R, Hocking J, Turner E, et al. 2018. An update on the RTTOV fast radiative transfer model (currently at version 12). *Geosci Model Dev*, 11(7): 2717-2937
- Schwander H, Kaifel A, Ruggaber A, et al. 2001. Spectral radiative-transfer modeling with minimized computation time by use of a neural-network technique. *Appl Opt*, 40(3): 331-335
- Singh R, Rayer P, Saunders R, et al. 2009. A fast radiative transfer model for the assimilation of water vapor radiances from the Kalpana very high resolution radiometer. *Geophys Res Lett*, 36(8): L08804
- Yu Z, Luo Y Z, Deng X F, et al. 2017. Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China using a spatially explicit machine learning algorithm. *Atmos Environ*, 155: 129-139
- Zhang P, Lu Q F, Hu X Q, et al. 2019. Latest progress of the Chinese meteorological satellite program and core data processing technologies. *Adv Atmos Sci*, 36(9): 1027-1045