基于地基辐射计模拟数据的大气温湿廓线 Sounding Transformer 反演方法

梁珍珍[1, 2]，郭伟[2,*]，王彩云[2]，刘鹏[2]，杨世杰[1, 2]，邢晴[1, 2]

1(中国科学院大学 北京 100049)

2(中国科学院国家空间科学中心 北京 100190)

摘　要　大气温湿廓线是数值天气预报和空间环境监测的关键参数。现有反演方法以线性回归和 BP 神经网络为主，前者难以刻画辐射传输非线性特征，后者对通道间关联表达不足。本文提出 Sounding Transformer 反演方法，利用自注意力机制学习通道间耦合关系，基于北京南郊观象台 2020-2025 年探空资料，用 MonoRTM 正演亮温构建数据集，实现温湿廓线反演。通过与线性回归、BP 神经网络对比，并开展噪声敏感性实验、消融实验、云液态水敏感性分析及 FY-3E VASS 产品比对校验，结果表明：（1）反演精度显著优于线性回归与 BP 神经网络，温度与湿度 RMSE 分别优化 22.4%、25.7%与 3.47%、2.07%；（2）加入系统噪声后优势依然保持；（3）时间周期编码、可降水量和液态水路径均为必要辅助特征；（4）K 波段对液态水敏感，V 波段氧气核心区基本不受影响。本研究为提高地基辐射计温湿廓线反演精度提供了方法支撑。

关键词　Transformer，温湿廓线反演，MonoRTM，自注意力机制，地基微波辐射计

中图分类号

最新录用

Sounding Transformer Retrieval Method for Atmospheric Temperature and Humidity Profiles Based on Ground-Based Radiometer Simulation Data

Zhenzhen Liang[1, 2]，Wei Guo[2,*]，Caiyun Wang[2]，Peng Liu[2]，Shijie Yang[1, 2]，Qing Xing[1, 2]

1 (University of Chinese Academy of Sciences, Beijing 100049)

2 (National Space Science Center, Chinese Academy of Sciences, Beijing 10019)

:

Abstract Atmospheric temperature and humidity profiles are key parameters for numerical weather prediction, space environment monitoring, and climate research. Ground-based microwave radiometers can provide continuous, all-weather observations of the troposphere, but their retrieval accuracy is fundamentally limited by the inversion algorithms employed. Existing retrieval methods are dominated by linear regression and backpropagation BP neural networks: linear regression struggles to characterize the strongly nonlinear features of radiative transfer in the microwave band, while BP neural networks inadequately express inter-channel correlations, since the brightness temperature channels of a radiometer are physically coupled through shared atmospheric absorption processes and thus exhibit significant redundancy and dependence. To address these limitations, this paper proposes a Sounding Transformer retrieval method that leverages the self-attention mechanism to learn coupling relationships among channels, capturing both global dependencies and nonlinear mappings from multi-channel brightness temperatures to atmospheric profiles. Based on radiosonde data from Beijing Nanjiao Observatory from 2020 to 2025, a simulated brightness temperature dataset is constructed using MonoRTM forward simulation, covering clear-sky and cloudy conditions. Temperature and humidity profiles are retrieved by incorporating auxiliary features such as time-period encoding, precipitable water vapor, and liquid water path. A comprehensive validation framework is established, including comparison with linear regression and BP neural networks, noise sensitivity experiments, ablation experiments, cloud liquid water sensitivity analysis, and validation against FY-3E VASS products. The results show that: (1) the retrieval accuracy is significantly superior to both baseline methods, with temperature and humidity RMSE improved by 22.4% and 25.7% relative to linear regression, and by 3.47% and 2.07% relative to BP neural networks; (2) the advantage remains stable after adding systematic noise, demonstrating robustness to radiometric calibration errors; (3) ablation results confirm that time-period encoding, precipitable water vapor, and liquid water path are all necessary auxiliary features; (4) K-band brightness temperatures are highly sensitive to liquid water content, while V-band oxygen absorption core channels are basically unaffected, providing physical interpretability of the attention weights; and (5) in comparison with FY-3E VASS satellite products, the temperature RMSE is 1.529 °C and absolute humidity RMSE is 0.359 g/m^3, with the most prominent difference in the near-surface layer, mainly attributable to satellite footprint averaging and surface emission effects. This study provides methodological support for improving the retrieval accuracy of temperature and humidity profiles from ground-based radiometers, and the proposed framework is extensible to other radiometric platforms and geographic stations.

Key words Transformer, temperature and humidity profile retrieval, MonoRTM, self-attention mechanism, ground-based microwave radiometer

最新录用

0 引言

大气温度和水汽廓线是表征大气热力状态、水汽输送和边界层结构演变的重要物理量，在数值天气预报、强对流天气监测、气候变化研究以及卫星遥感产品验证中均具有重要作用**错误!未找到引用源。**。常规探空能够提供精度较高的温湿廓线资料，但受观测时次和站点分布限制，通常只能在固定时刻施放，难以满足边界层快速变化过程和连续业务应用对高时间分辨率资料的需求**错误!未找到引用源。**。因此，发展可连续获取大气温湿廓线的地基遥感反演方法，是提升边界层大气监测能力的重要途径。

现有大气温湿廓线反演方法主要包括物理反演、统计回归和机器学习方法。物理反演基于大气辐射传输方程，物理机理清晰，但反演过程高度非线性，计算量较大，在实时业务应用中受到限制。统计回归方法通过建立亮温与大气参数之间的经验关系实现快速反演，计算效率高、实现简单，但对辐射传输过程中的复杂非线性特征刻画不足，在湿度廓线和复杂天气条件下精度有限。随着机器学习方法的发展，反向传播神经网络**错误!未找到引用源。**（Back Propagation Neural Network, BPNN）和卷积神经网络（Convolutional Neural Network, CNN）已被广泛用于微波辐射计廓线反演**错误!未找到引用源。**，相比传统线性方法已有明显改进**错误!未找到引用源。错误!未找到引用源。**。然而，BP 和 CNN 网络结构对通道间谱相关关系和廓线垂直层次关联的表达不足**错误!未找到引用源。错误!未找到引用源。**。

Transformer 基于自注意力机制**错误!未找到引用源。**，能够在输入序列内部建立全局依赖关系。鉴于辐射计多通道之间存在复杂的非线性耦合关系，Transformer 可利用自注意力机制学习通道间相互作用，进而提高温湿廓线反演精度。但原始 Transformer 主要面向自然语言处理中的离散序列建模**错误!未找到引用源。**，其输入表示、位置编码和输出结构不能直接套用于连续物理量驱动的廓线回归任务**错误!未找到引用源。错误!未找到引用源。**，因此需要结合地基微波辐射计观测特点进行改造。

针对上述问题，本文利用 Transformer 的自注意力机制，提出一种 Sounding Transformer 反演方法。以中国科学院国家空间科学中心研制的 46 通道 K/V 双波段地基辐射计为仪器原型，基于北京南郊观象台 2020—2025 年探空资料，利用 MonoRTM（Monochromatic Radiative Transfer Model）辐射传输模型正演构建模拟亮温数据集。将 46 通道亮温视为具有物理含义的序列输入，联合时间周期编码、可降水量和液态水路径等辅助特征，利用多头自注意力机制学习 K/V 波段通道间耦合关系，实现 93 层温度与绝对湿度廓线的联合反演。通过与线性回归、BP 神经网络对比，并开展噪声对反演精度影响实验、扩展消融实验、云液态水含量敏感性分析及 FY-3E VASS (Vertical Atmospheric Sounding System)产品比对校验实验，综合评估模型性能。旨在为提高地基辐射计的温湿廓线反演精度，进而提升数值天气预报准确性与空间环境变化监测能力提供方法支撑。

1. 基于 MonoRTM 的大气温湿廓线正演

1.1 地基辐射计仪器原型

仪器原型为中国科学院国家空间科学中心 2026 年研制的基于 SDR 的狄克型 K/V 双波段地基辐射计，工作于 K 波段（21.3–31.7GHz，水汽通道）和 V 波段（50.8–61.2GHz，氧气通道），主要技术指标汇总于表 1。

表 1　基于 SDR 的 K/V 双波段狄克型辐射计技术指标

Table 1　Technical Specifications of SDR-Based K/V Dual-Band Dicke Radiometer

参数	技术指标
频率	K：21.3~31.7 GHz；V：50.8~61.2 GHz
通道数	K：23；V：23

带宽	288 MHz (1~4 通道,19~23 通道)；400 MHz (5~18 通道)
系统稳定性	≤1.0 K/年
亮温测量范围	0-400 K
灵敏度	K：0.15K；V：0.21K
积分时间	1s
通道间一致性	≤1K
天线增益	25dBi
天线驻波比	1.25
极化方式	线极化
波束宽度	5-7°
高温源	70.2℃

K 波段和 V 波段各配置 23 个通道，共 46 个通道。K 波段 23 个通道的中心频率依次为：21.46、21.82、22.18、22.54、23.00、23.50、24.00、24.50、25.00、25.50、26.00、26.50、27.00、27.50、28.00、28.50、29.00、29.50、30.00、30.46、30.82、31.18、31.54GHz；V 波段 23 个通道的中心频率依次为：50.96、51.32、51.68、52.04、52.50、53.00、53.50、54.00、54.50、55.00、55.50、56.00、56.50、57.00、57.50、58.00、58.50、59.00、59.50、59.96、60.32、60.68、61.04GHz。

1.2 数据来源及预处理

北京南郊观象台（世界气象组织（World Meteorological Organization, WMO）站号 54511，39.93°N，116.28°E，海拔 32 m）位于华北平原北部，地处城市下垫面与郊区的过渡带，属全球探空观测网（Global Climate Observing System Upper-Air Network, GUAN）核心站点，每日在北京时间 08 时和 20 时施放探空气球，观测要素包括气压、温度、相对湿度、风速风向及位势高度等，原始探空资料经业务质量控制后，用于辐射传输正演与模型训练。

本研究使用 54511 站 2020 年 1 月至 2025 年 12 月的探空资料开展模拟地基微波辐射计反演实验。经质量控制后，共获得 3713 个有效样本，各年份样本数较为均衡（2020-2025 年分别为 651、628、610、601、605、618 份）。天气类型依据 Open-Meteo API **错误!未找到引用源。**提供的 ERA5 再分析资料，并参考 WMO 天气编码**错误!未找到引用源。**进行判定，其中晴天样本 3430 个（92.4%），云天/降水样本 283 个（7.6%）。样本按 8:2 划分为训练集与独立测试集，其中训练样本 2970 个，测试样本 743 个。样本在年份、季节、时次和天气条件上的分布见图 1。

最终版用

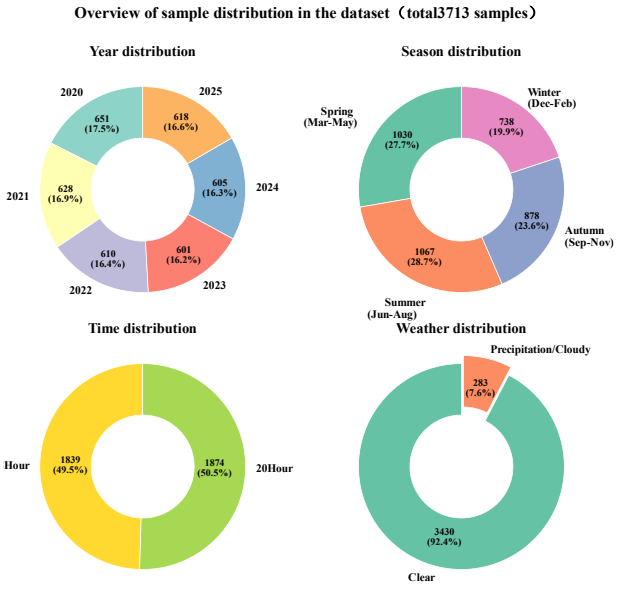


图 1 数据集样本在年份、季节、时间和天气类型上的分布

Fig.1 Dataset sample distribution across year, season, time, and weather type

原始探空廓线的垂直层次存在差异，本文将其统一插值至 93 层标准高度网格。该网格覆盖近地面至 10–12 km 高度，呈近密远疏分布：0–500m 为 10–30m，500–2000m 为 40–90m，2000–12000m 为 100–300m。插值时，0–500m 近地面层采用三次样条方法，500m 以上中高层采用分段线性方法，以尽量保留边界层温湿结构。本文进一步开展的质量控制主要包括：检查气压、高度、温度、露点温度和相对湿度等必备要素，剔除缺测值、重复层和明显异常层，并要求有效层数不少于 4 层且最大探测高度不低于 10km。处理后的温湿廓线统一保存为 NetCDF4 格式，作为 MonoRTM 正演和模型训练输入。

1.3 MonoRTM 辐射传输模型及辐射亮温正演

MonoRTM 由美国大气环境研究室（Atmospheric and Environmental Research, AER）开发，分子吸收参数主要来自 HITRAN 数据库**错误!未找到引用源。**，包括水汽压力加宽系数、温度加宽系数等。水汽连续吸收采用 MT_CKD 模型**错误!未找到引用源。**，MonoRTM 同时考虑了氧气、二氧化碳、臭氧和氮等气体的压力加宽与自加宽效应。液态水吸收基于 Liebe 介电模型计算**错误!未找到引用源。**。MonoRTM 模型提供多种辐射传输方案，本文选用 case3（用户提供大气廓线）和 case5（含云条件模拟），以适应晴空和含云两类天气场景。

正演前，根据地面天气状况将样本划分为晴空和含云两类。晴空模式：MonoRTM 直接读取 93 层大气温湿廓线，读取 TAPE3 格式的分子吸收线参数，逐频率计算晴空大气的吸收系数与光学厚度，不考虑液态水吸收。含云模式采用两步计算：第一步以晴空模式运行 MonoRTM 并输出参考光学厚度文件；第二步由探空相对湿度廓线利用公式（1）计算液态水含量（Liquid Water Content, LWC），式中，RH为相对湿度，进而将 LWC 廓线作为输入重新运行含云条件下的 MonoRTM 正演，得到含云场景的模拟亮温。

$$LWC(\rho)=\begin{cases}0 & \text{if } RH<85\% \\ \dfrac{2.0\times(RH-85)}{10}\,\mathrm{g/m^3} & \text{if } 85\%\le RH\le 95\% \\ 2.0\mathrm{g/m^3} & \text{if } RH>95\%\end{cases} \tag{1}$$

正演后，各样本得到 46 个频率通道的模拟亮温、可降水量（Precipitable Water Vapor, PWV）、大气透过率、液态水路径（Liquid Water Path, LWP）等物理参量。

从正演结果中提取 46 维亮温向量，与 6 维辅助特征共同构成模型输入。辅助特征包括

年积日周期编码（$d_{\sin}$、$d_{\cos}$）、UTC 时刻周期编码（$h_{\sin}$、$h_{\cos}$）、PWV 和 LWP。其中，$d_{\sin}=\sin(2\pi\cdot d/365.25)$、$d_{\cos}=\cos(2\pi\cdot d/365.25)$、$h_{\sin}=\sin(2\pi\cdot h/24)$、$h_{\cos}=\cos(2\pi\cdot h/24)$，d 为年积日，h 为 UTC 时刻。模型输出为 93 层温度和绝对湿度，其中温度单位为℃，绝对湿度单位为$\mathrm{g/m^3}$。

1.4 亮温观测噪声模拟

相较于实际观测值，辐射传输模型正演得到的亮温通常不包含仪器噪声、定标误差及观测环境扰动等误差来源。为使其更接近实际观测条件，同时检验模型在存在观测误差条件下的稳定性与泛化能力，本文在亮温输入端引入高斯随机噪声。对每一样本的 46 个通道亮温分别加入均值为 0、标准差为 σ 的高斯噪声，即公式（2），

$$T_b^{noise}=T_b+\varepsilon,\quad \varepsilon\sim N(0,\sigma^2) \tag{2}$$

式中，T_b为前文正演得到的原始亮温，T_b^{noise}为加噪后的模拟亮温。

2. 基于 Transformer 模型的大气温湿廓线反演

如前文所述，地基微波辐射计各通道亮温并非独立观测，通道间存在相关性和耦合性。本文将多通道亮温作为具有物理含义的序列输入，利用 Transformer 自注意力机制学习频率通道间的耦合关系，联合输出 93 层温湿廓线。

本节对标准 Transformer 结构进行改造，构建由输入层、编码器和输出层组成的 Sounding Transformer 模型：去除用于序列生成的 Decoder，保留编码器中的多头自注意力（Multi-Head Self-Attention, MHSA）和前馈网络（Feed-Forward Network, FFN）子层，并设置独立的温度和绝对湿度输出头。基于前述加噪后的模拟亮温数据集开展模型训练。

2.1 模型架构

Sounding Transformer 模型架构如图 2 所示。将 46 个亮温 token 和 1 个辅助信息 token（6 维辅助特征向量）拼接成长度为 47 的输入序列。其中，各通道亮温经逐通道线性投影得到$\mathrm{h}_i=\mathrm{W}_i\mathrm{T}_b^i+\mathrm{b}_i\ (i=1,\dots,46)$，式中，$\mathrm{W}_i\in\mathrm{R}^{d_{model}}$、$\mathrm{b}_i\in\mathrm{R}^{d_{model}}$分别为第$i$通道的投影权重向量与偏置向量。辅助信息经线性投影得到：$\mathrm{h}_{aux}=\mathrm{W}_{aux}\mathrm{x}_{aux}+\mathrm{b}_{aux}$，式中，$\mathrm{x}_{aux}\in\mathrm{R}^6$为辅助信息特征向量，$\mathrm{W}_{aux}\in\mathrm{R}^{d_{model}\times 6}$、$\mathrm{b}_{aux}\in\mathrm{R}^{d_{model}}$分别为投影权重矩阵与偏置向量。最终得到输入序列$\mathrm{H}^{(0)}=[\mathrm{h}_1,\dots,\mathrm{h}_{46},\mathrm{h}_{aux}]^T\in\mathrm{R}^{47\times d_{model}}$，式中$\mathrm{h}_1,\dots,\mathrm{h}_{46}$为各通道亮温经过嵌入后的特征向量，$d_{model}$=64 为模型嵌入维度。这种设计使模型同时学习通道间关系与大气背景约束。亮温 token 采用逐通道线性投影，以保留不同频率通道的独立物理含义；辅助信息 token 采用两层全连接映射，以增强辅助特征的非线性表达能力。与将所有变量直接拼接输入全连接网络相比，这种设计更符合多通道亮温的序列特征，也为后续自注意力建模提供了结构化的输入。

在位置编码中，考虑到通道间亮温的耦合关系，模型采用可学习的位置编码**错误!未找到引用源。**：$\mathrm{H}^{(0)}\leftarrow\mathrm{H}^{(0)}+\mathrm{E}_{pos}$，其中$\mathrm{E}_{pos}\in\mathrm{R}^{47\times d_{model}}$为可学习位置编码矩阵。与固定正弦编码相比，可学习位置编码可根据训练数据自适应刻画频率通道之间的物理关联。输入序列经层归一化（LayerNorm）和正则化**错误!未找到引用源。**（Dropout）处理，再送入编码器。

编码器由 2 个编码器子层堆叠而成。每个编码器层均包含 MHSA 和 FFN 两个子模块，并采用 Pre-Norm **错误!未找到引用源。**残差连接，以提高训练稳定性。MHSA 的作用机制为：给定输入$\mathrm{X}\in\mathrm{R}^{n\times d}$（n 为序列长度，d 为特征维度），通过线性变换生成查询（Query）、键（Key）和值（Value）矩阵：$\mathrm{Q}=\mathrm{XW}_i^Q$，$\mathrm{K}=\mathrm{XW}_i^K$，$\mathrm{V}=\mathrm{XW}_i^V$，其中$\mathrm{W}_i^Q$，$\mathrm{W}_i^K$，$\mathrm{W}_i^V$为第$i$个注意力头的投影矩阵。将特征空间划分为 h 个子空间并行计算多头注意力，h 为注意力头总数。注意力计算过程如式（3）和式（4）所示，W^O为输出投影矩阵。

$$\mathrm{MultiHead(X)}=\mathrm{Concat(head}_1,\dots,\mathrm{head}_h)\mathrm{W}^O \tag{3}$$

$$\mathrm{head}_i = \mathrm{Attention}\left(\mathrm{XW}_i^Q, \mathrm{XW}_i^K, \mathrm{XW}_i^V\right), i = 1, \ldots, h \tag{4}$$

本文设置嵌入维度为 64，注意力头数为 4。不同注意力头可从不同子空间捕获 K 波段水汽吸收、V 波段氧气吸收以及跨频段关系。MHSA 子层通过 Pre-Norm 残差连接：

$$\mathrm{H}' = \mathrm{H}^{(l-1)} + \mathrm{Dropout}\left(\mathrm{MultiHead}\left(\mathrm{LayerNorm}\left(\mathrm{H}^{(l-1)}\right)\right)\right) \tag{5}$$

FFN 对每个 Token 独立施加两层非线性变换：$\mathrm{FFN(x)} = \mathrm{GELU(xW_1 + b_1)W_2 + b_2}$，其中x为单个 token 的特征向量，$\mathrm{W_1} \in \mathrm{R}^{d_{model} \times d_{ff}}$，$\mathrm{W_2} \in \mathrm{R}^{d_{ff} \times d_{model}}$，$\mathrm{b_1} \in \mathrm{R}^{d_{ff}}$，$\mathrm{b_2} \in \mathrm{R}^{d_{model}}$，$d_{ff} = 4d_{model}$。激活函数采用 GELU：$\mathrm{GELU}(x) = x \cdot \Phi(x) = x \cdot \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$，以增强模型对亮温—廓线非线性映射的表达能力。FFN 子层同样采用 Pre-Norm 残差连接，经2层编码器处理后，输出包含通道关联信息的高维序列表示。

编码器输出经均值池化与 MLP 降维，将序列信息聚合为固定维度特征，如式（6）（7）所示。

$$\mathrm{f}_{flat} = \mathrm{MeanPool}\left(\mathrm{H}^{(N_L)}\right) \in \mathrm{R}^{47 \times d_{model}} \tag{6}$$

$$\mathrm{f} = \mathrm{MLP}_{pool}\left(\mathrm{f}_{flat}\right) \in \mathrm{R}^{2 \times d_{model}} \tag{7}$$

池化特征通过独立输出头分别预测 93 层温度与 93 层绝对湿度，如式（8）和式（9）所示。温度预测头包含两层：$\mathrm{W}_{T1} \in \mathrm{R}^{128 \times 128}$、$\mathrm{b}_{T1} \in \mathrm{R}^{128}$为第一层参数，$\mathrm{W}_{T2} \in \mathrm{R}^{128 \times 128}$、$\mathrm{b}_{T2} \in \mathrm{R}^{128}$为第二层参数；温度预测头结构相同，参数记为$\mathrm{W}_{\rho 1}$、$\mathrm{b}_{\rho 1}$、$\mathrm{W}_{\rho 2}$、$\mathrm{b}_{\rho 2}$。

$$\hat{\mathrm{T}} = \mathrm{W}_{T2} \cdot \mathrm{GELU}(\mathrm{W}_{T1}\mathrm{f} + \mathrm{b}_{T1}) + \mathrm{b}_{T2} \in \mathrm{R}^{93} \tag{8}$$

$$\hat{\rho}_w = \mathrm{W}_{\rho 2} \cdot \mathrm{GELU}\left(\mathrm{W}_{\rho 1}\mathrm{f} + \mathrm{b}_{\rho 1}\right) + \mathrm{b}_{\rho 2} \in \mathrm{R}^{93} \tag{9}$$

温度和绝对湿度分支共享同一编码器，以利用二者同源于大气热力状态的物理联系；采用独立输出头分别解码，避免两个任务在输出阶段相互干扰。

最新学用

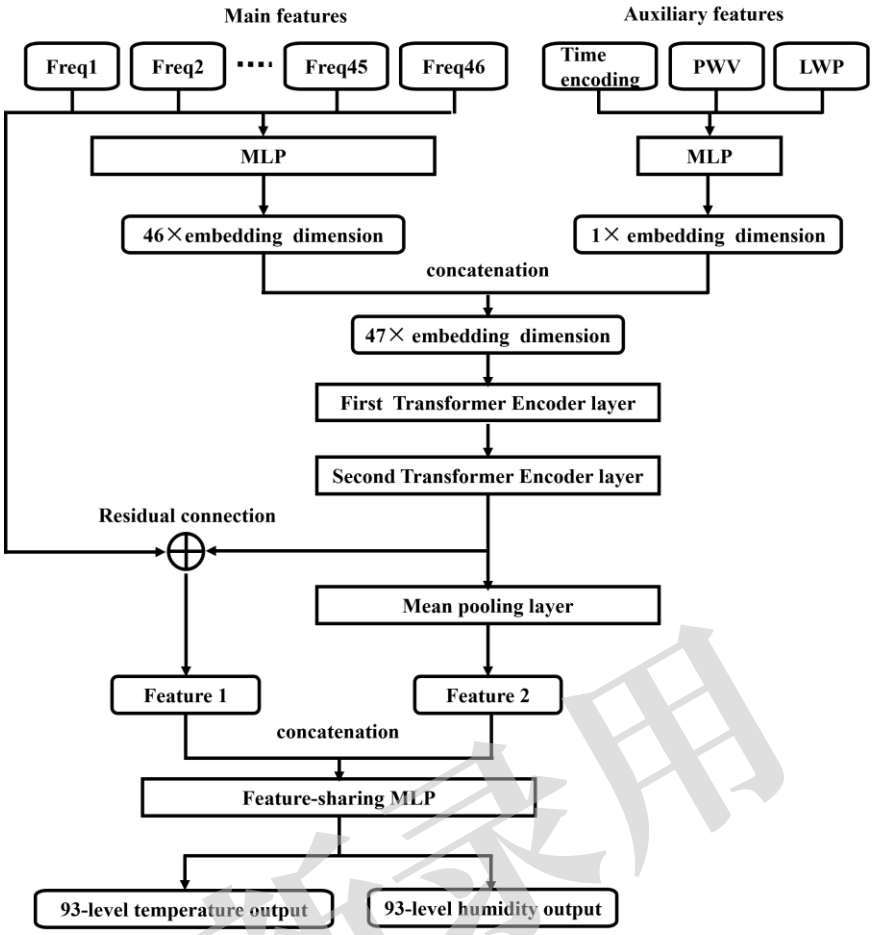


图 2 Sounding Transformer 模型总体架构

Fig.2 Overall architecture of the Sounding Transformer model

2.2 损失函数

模型采用加权联合均方误差（Mean Squared Error, MSE）**错误!未找到引用源。**作为损失函数，开展端到端训练。

$$\mathcal{L} = \lambda_T \cdot \frac{1}{N \cdot L} \sum_{i=1}^{N} \sum_{l=1}^{L} (\hat{T}_{il}^{(s)} - T_{il}^{(s)})^2 + \lambda_\rho \cdot \frac{1}{N \cdot L} \sum_{i=1}^{N} \sum_{l=1}^{L} (\hat{\rho}_{w,il}^{(s)} - \rho_{w,il}^{(s)})^2 \tag{10}$$

其中，N 为样本数，L 为高度层数，λ_T与λ_ρ分别为温度与绝对湿度的损失权重。引入可调权重的原因是：标准化后温度误差与湿度误差的收敛速率存在差异，等权重处理易导致某一任务主导优化方向。通过调节λ_T与λ_ρ的比值可平衡多任务优化目标。

训练前，对所有输入特征和目标变量进行标准化**错误!未找到引用源。**。模型权重采用 Xavier 均匀初始化**错误!未找到引用源。**。优化器选用 AdamW，初始学习率为 5×10^{-4}，权重衰减系数为 1×10^{-3}，批量大小为 64，训练轮次为 300。学习率采用余弦退火策略**错误!未找到引用源。**，Dropout 概率为 0.15，梯度裁剪阈值为 1.0，早停耐心为 45 个 epoch **错误!未找到引用源。**。

2.3 模型训练

图 3 展示了 σ=0.5K 噪声条件下，前 175 个训练轮次内损失与均方根误差（Root Mean Square Error, RMSE）、决定系数（coefficient of determination，R^2）的变化。训练损失与验证损失在前 25 个 epoch 快速下降。第 51、151 轮附近的波动对应学习率暖重启（warm restart）。随训练推进，温度 RMSE 从 3.934°C 降至 1.702°C，R^2从 0.961 提升至 0.9926；绝对湿度 RMSE 从 1.920g/m³降至 0.854g/m³，R^2从 0.8404 提升至 0.9684。

最新录用

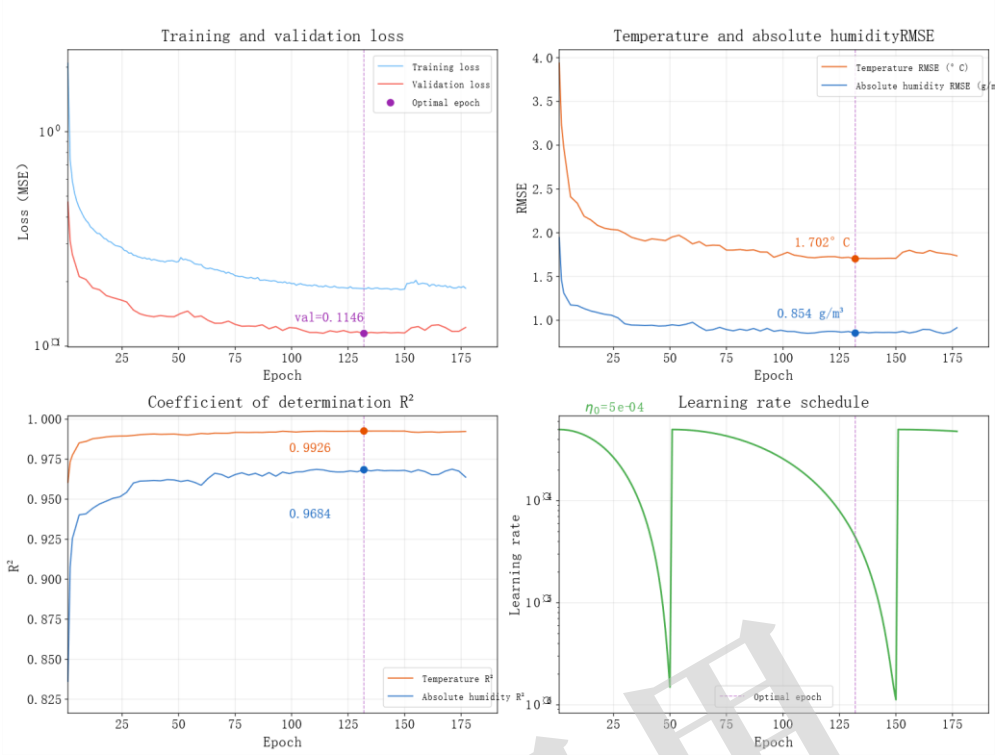


图 3 训练过程中损失、温度与绝对湿度 RMSE、决定系数 R^2及学习率变化（σ=0.5K）

Fig.3 Variation of loss, RMSE, coefficient of determination R^2, and learning rate during training (σ=0.5K)

图 4 给出了 σ=0.5K 条件下不同高度层 RMSE 随训练轮次的变化。温度反演中，0～2km 高度层收敛最快，最终 RMSE 值为 1.07℃；10km 以上高层残余误差最大，最终 RMSE 值为 3.24℃；绝对湿度反演误差随高度升高而减小，低层 RMSE 为 1.061g/m^3，高层 RMSE 为 0.050g/m^3，这与水汽主要集中在对流层低层的物理特性一致。

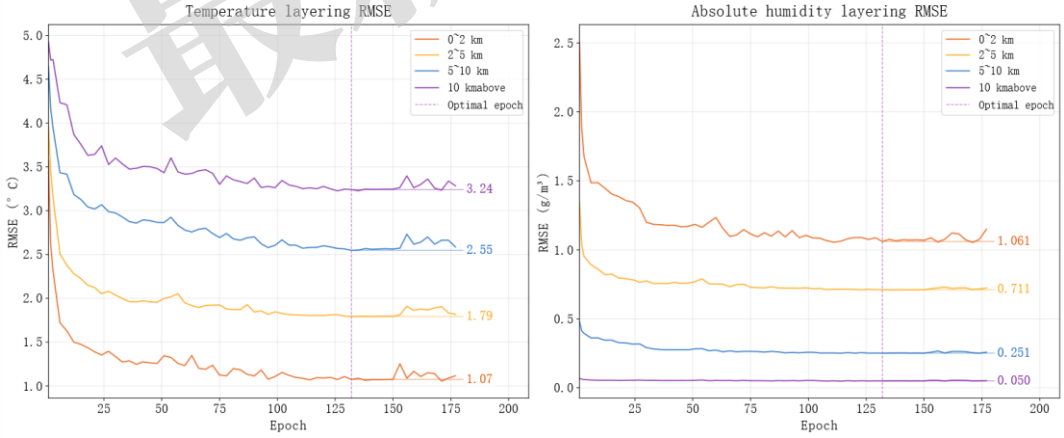


图 4 不同高度层温度 RMSE（左）和绝对湿度 RMSE（右）随训练轮次的变化（σ=0.5K）

Fig.4 Variation of temperature RMSE (left) and absolute humidity RMSE (right) with training epochs at four altitude levels (σ=0.5K)

2.4 对比基线方法

线性回归对比基线选用岭回归：采用 L2 正则化**错误!未找到引用源。**的线性模型，参数量约 0.99 万。岭回归可通过解析解快速求得参数，计算效率较高，但难以描述亮温与温湿廓线之间的非线性关系。

BP 神经网络：采用三层全连接结构，其中输入 52 维（46 通道亮温+6 维辅助特征）、隐

含 256 个神经元、输出 186 维（93 层温度+93 层湿度），隐含层使用 Tanh 激活函数，参数量约 6.1 万。隐含层维度参考张雪慧等**错误!未找到引用源。**的经验公式确定。训练策略与 Transformer 保持一致，确保对比的公平性。

3. 结果与讨论

3.1 反演精度对比

所有实验基于 PyTorch 框架**错误!未找到引用源。**实现，采用相同的数据划分。表 2 从 RMSE、平均绝对误差（Mean Absolute Error, MAE）和R^2三项指标对比了不同方法在 σ=0.5K 条件下的反演精度。

表 2 不同反演方法的温度与绝对湿度性能比较(σ=0.5K)

Table 2 Performance comparison of different retrieval methods for temperature and absolute humidity (σ=0.5K)

方法	温度 RMSE	温度 MAE	温度R^2	湿度 RMSE	湿度 MAE	湿度R^2
岭回归	2.188	1.392	0.9878	1.142	0.688	0.9435
BP 网络	1.759	1.245	0.9921	0.866	0.556	0.9676
Transformer	**1.702**	**1.225**	**0.9926**	**0.854**	**0.556**	**0.9684**

由表 2 可见，随着模型非线性表达能力和通道关联建模能力的增强，反演精度逐步提高。与岭回归相比，Transformer 的温度 RMSE 降低 22.4%，绝对湿度 RMSE 降低 25.7%；与 BP 神经网络相比，温度 RMSE 与绝对湿度 RMSE 分别降低 3.47%和 2.07%。Transformer 相较于 BP 网络在温度反演中的优势更为明显。这主要是因为温度廓线垂直变化相对平滑且结构稳定，V 波段亮温与温度廓线之间的统计映射规律性较强，自注意力机制能够更好地捕捉并利用通道间的关联信息来约束温度廓线。相比之下，绝对湿度廓线垂直变化剧烈，受液态水含量变化及边界层混合过程的影响，K 波段亮温与湿度廓线之间的统计映射更为复杂。因此，BP 神经网络已能提取大部分有效信息，Transformer 带来的绝对湿度反演提升幅度小于温度反演。

3.2 噪声对反演精度影响分析

为了分析噪声对反演精度影响，本文开展了不同噪声水平下三种方法的反演精度对比实验，结果如图 5 所示。在 0–1.75K 噪声水平范围内，Transformer 在温度和绝对湿度反演中均保持最优，说明 Sounding Transformer 模型对仪器随机误差具有较好的鲁棒性。当噪声水平增大至 2.0K 及以上时，各方法误差均明显增大；Transformer 的温度反演仍优于 BP 网络和岭回归，但湿度指标与 BP 网络趋于接近。总体来看，Transformer 在接近真实观测噪声（本研究原型辐射计灵敏度 K 波段为 0.15K，V 波段为 0.21K）时优势稳定。本文后续图表均以 σ=0.5K 为例进行分析。

最新录用

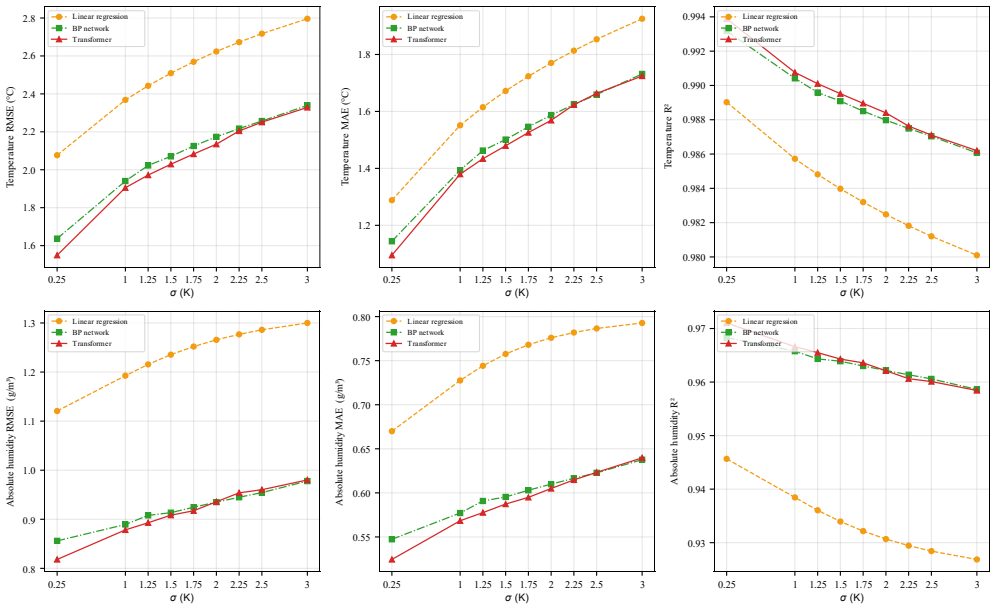


图 5 不同高斯噪声水平下三种方法温度（上排）与绝对湿度（下排）RMSE、MAE 及 R^2变化

Fig.5 Variation of RMSE, MAE and R^2 for temperature (top row) and absolute humidity (bottom row) among three methods under different Gaussian noise levels

3.3 误差垂直分布特征

（1）整体垂直误差特征

图 6 给出三种反演方法误差指标的垂直分布。温度反演方面，RMSE 随高度升高而增大：0–1km 高度层三种方法误差均较低，其中 Transformer 最优（RMSE 约 0.7–1°C，MAE 约 0.6–0.7°C）；3–8km 中层误差差距逐渐拉开，线性回归 RMSE 达到 2.2–3.0°C，Transformer 在 1.7–2.4°C；10km 附近各方法 RMSE 均较大，线性回归超过 3.8°C，Transformer 控制在 3.1°C。绝对湿度反演误差随高度升高而减小，近地面 RMSE 与 MAE 均最大（分别约为 0.9–1.4g/m^{3}和 0.7–1.05g/m^3），随高度迅速减小，5km 以上 RMSE 降至 0.4g/m^{3}以下，10km 处接近 0.1g/m^3，这主要是因为水汽主要集中在对流层低层。总体来看，Transformer 在全高度范围内均保持较低误差，尤其在 3–10km 相对于传统方法的改善更为明显。

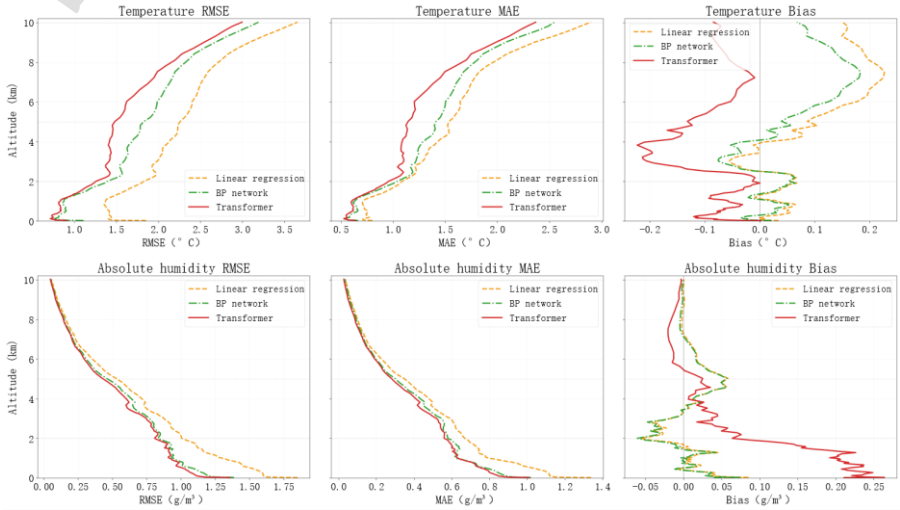


图 6 三种反演方法温度（上排）与绝对湿度（下排）RMSE、MAE 垂直分布对比

Fig.6 Vertical distribution comparison of RMSE and MAE for temperature (top row) and absolute humidity (bottom row) among three inversion methods

（2）决定系数的垂直变化

图 7 给出了R^2的垂直分布。温度 R^2在近地面接近 0.99，随高度升高逐渐下降；10km 附近 Transformer 约为 0.865，BP 网络约为 0.85，线性回归约为 0.8。绝对湿度 R^2在 0–2km 较高，随高度下降较快，在 5–7km 出现短暂的平台期后继续下降；10km 附近 Transformer 约为 0.7，BP 网络约为 0.65，线性回归约为 0.58。需要指出，高层湿度 R^2下降并不意味着绝对误差较大。原因在于高层水汽含量很低且样本方差较小，少量绝对偏差即可导致R^2显著下降。因此，评价高层湿度反演性能时应综合参考 RMSE 等绝对误差指标。

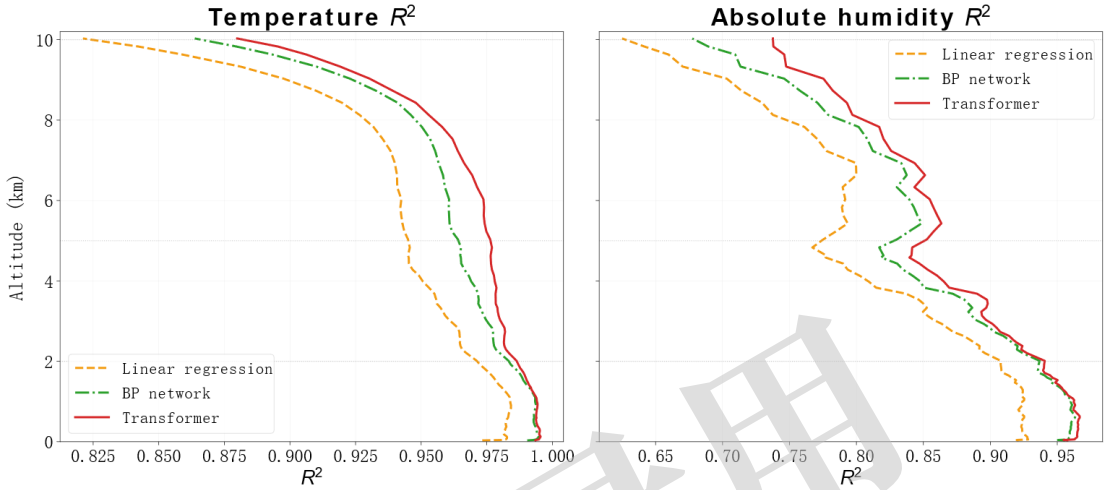


图 7 三种方法的温度（左）和绝对湿度（右）决定系数R^2垂直分布

Fig.7 Vertical distribution of coefficient of determination R^2 for temperature (left) and absolute humidity (right) among three methods

3.4 扩展消融实验

为检验模型输入特征和结构设置的合理性，本文开展扩展消融实验。从辅助特征和模型深度两个方面分析各因素对反演精度的影响。实验结果如表 3 所示。

在辅助特征方面，去除全部辅助特征后温度 RMSE 由 1.702°C 增至 2.447°C，湿度 RMSE 由 0.854g/m³增至 1.112g/m³。单独去除时间特征后温度 RMSE 增至 2.447°C，湿度 RMSE 增至 1.108g/m³。在所有单项辅助特征中，时间特征影响最为显著，这表明时间编码对温湿反演约束最强。去除 PWV 或 LWP 后湿度 RMSE 均升至约 0.90g/m³，温度 RMSE 升至约 1.91°C，这说明二者对温度和湿度反演均有补充贡献。进一步分析表明，去除 LWP 后温度 RMSE 相对全模型增大 11.93%，湿度 RMSE 增大了 5.76%，云属性对温湿度廓线的反演有不可忽视的贡献。

在模型深度方面，1 层和 6 层编码器的温度 RMSE 分别为 1.922°C 和 1.876°C，均高于 2 层的 1.702°C，表明当前数据规模下 2 层编码器已达到较优的拟合效果，层数增加并未带来显著受益。

表 3 扩展消融实验结果（σ=0.5K）

Table 3 Results of extended ablation experiments (σ=0.5K)

实验组	温度 RMSE/°C	温度 MAE/°C	温度 R^2	湿度 RMSE/(g/m³)	湿度 MAE/(g/m³)	湿度 R^2
完整模型	1.702	1.225	0.996	0.854	0.556	0.968 4

录用稿

仅亮温	2.447	1.563	0.9848	1.112	0.632	0.9464
无时间特征	2.447	1.553	0.9848	1.108	0.617	0.9469
无 PWV	1.913	1.389	0.9907	0.902	0.592	0.9648
无 LWP	1.905	1.387	0.9908	0.897	0.588	0.9652
无 PWV、LWP	1.929	1.400	0.9905	0.909	0.598	0.9642
1 层	1.922	1.400	0.9906	0.880	0.574	0.9665
6 层	1.876	1.360	0.9910	0.895	0.570	0.9654

3.5 云液态水含量敏感性分析

为分析云属性对通道亮温模拟的敏感性，本文从有云样本中选取 5 个代表性样本，覆盖低、中、高不同 LWP 水平；每个样本分别采用晴空方案和含云方案（含 LWC）进行 MonoRTM 正演；并计算两种方案通道亮温的差值，即 ΔTB=TB 含云-TB 晴空。实验结果如图 8 所示，K 波段 21–31GHz 对液态水含量非常敏感，随 LWP 增大，亮温增量显著上升；V 波段氧气吸收核心区敏感性较低，其中 50–54GHz 过渡频段仍受液态水连续吸收影响，但 55–61GHz 氧气强吸收通道基本不受液态水含量的影响。

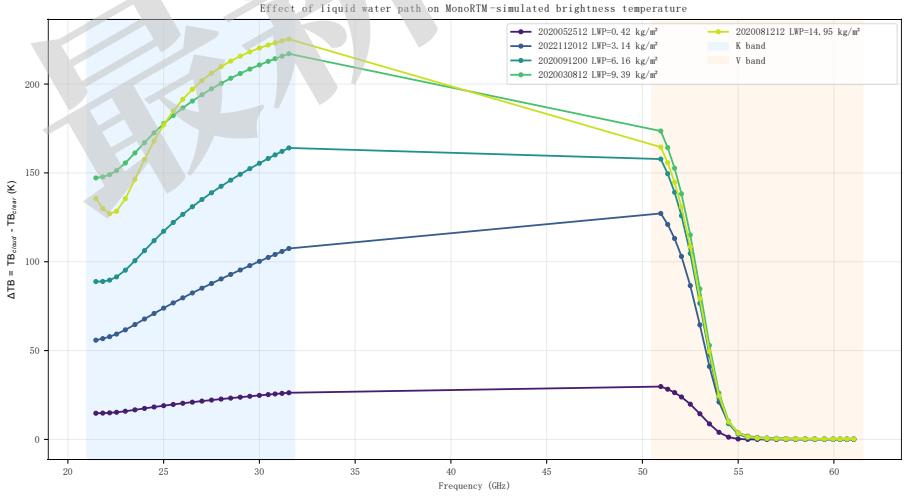


图 8 云的属性对通道亮温模拟的敏感性分析

Fig.8 Sensitivity Analysis of Cloud Properties on Channel Brightness Temperature Simulation

3.6 与 FY-3E VASS 大气廓线产品比对校验

本文采用 FY-3E VASS 2025 年全年大气廓线产品，开展比对校验实验，数据来自国家卫星气象中心风云卫星遥感数据服务网。按照时空匹配原则进行匹配，时间窗口为探空时刻±3h 内，匹配空间为 54511 站点经纬度的±2°范围内。最终选取北京 54511 站附近升轨样本 354 个、降轨样本 342 个，共 696 个时空匹配样本。统计时，将 VASS 廓线和 Transformer 反演廓线统一插值到探空标准高度层，并在 0–2km、2–5km、5–8km 和 8–12km 高度范围内分别

最新录用

计算其与探空真值之间的 RMSE。升轨时间对应北京时间 8：00 前后，降轨时间对应北京时间 20：00 前后。升轨产品与 Transformer 反演廓线和探空真值的 RMSE 比对结果如图 9 所示，降轨产品与 Transformer 反演廓线和探空真值的 RMSE 比对结果如图 10 所示。

需要说明的是，Transformer 反演基于 MonoRTM 正演亮温，输入为垂直天顶方向的地基观测；FY-3E VASS 为实际星载反演产品，受斜向观测几何、云污染、地表发射率等因素影响，两者反演条件存在本质差异。因此，以下 RMSE 对比仅作为参照校验，不代表地基辐射计精度优于星载仪器。

总体来看，Transformer 反演结果温度 RMSE 低于 VASS 产品，且离散度较小；降轨时近地层差异最为明显。湿度方面，Transformer 在 0–5km 层绝对湿度 RMSE 低于 VASS 产品；随着高度升高，绝对湿度量级减小，两者 RMSE 均较低。结果表明，Transformer 反演在不同轨道方向和不同高度层上均具有较好的稳定性。

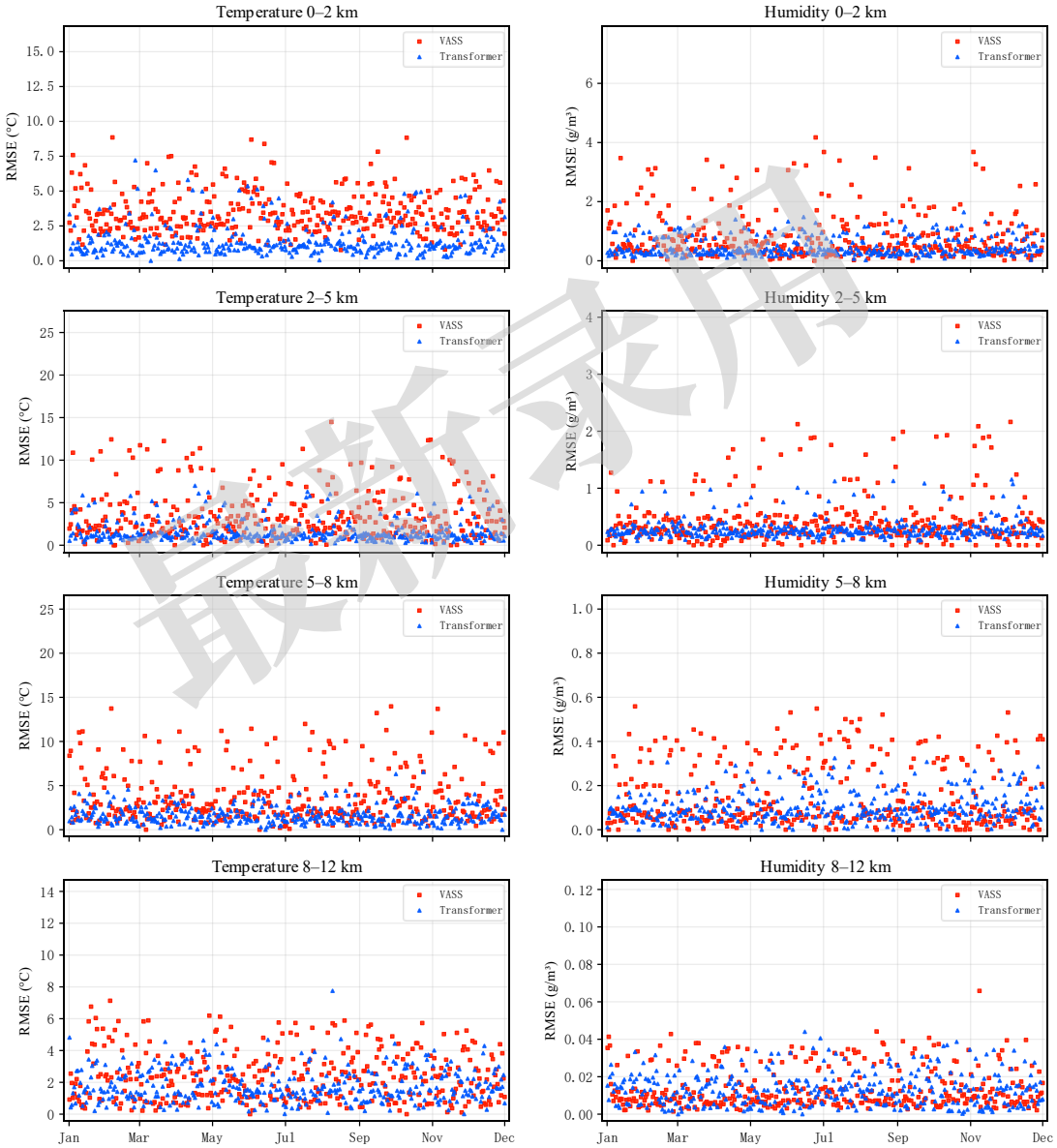


图 9 2025 年 Transformer 反演、FY-3E VASS 产品（升轨）与探空真值之间各高度层的 RMSE 对比（左列：温度，右列：绝对湿度，从上到下依次为：0~2 km、2~5 km、5~8 km、8~12 km）

Fig.9 RMSE comparison among Transformer retrieval, FY-3E VASS product (ascending orbit), and radiosonde truth at different altitude levels in 2025 (left column: temperature, right column: absolute humidity, top to bottom: 0~2 km、2~5 km、5~8 km、8~12 km)

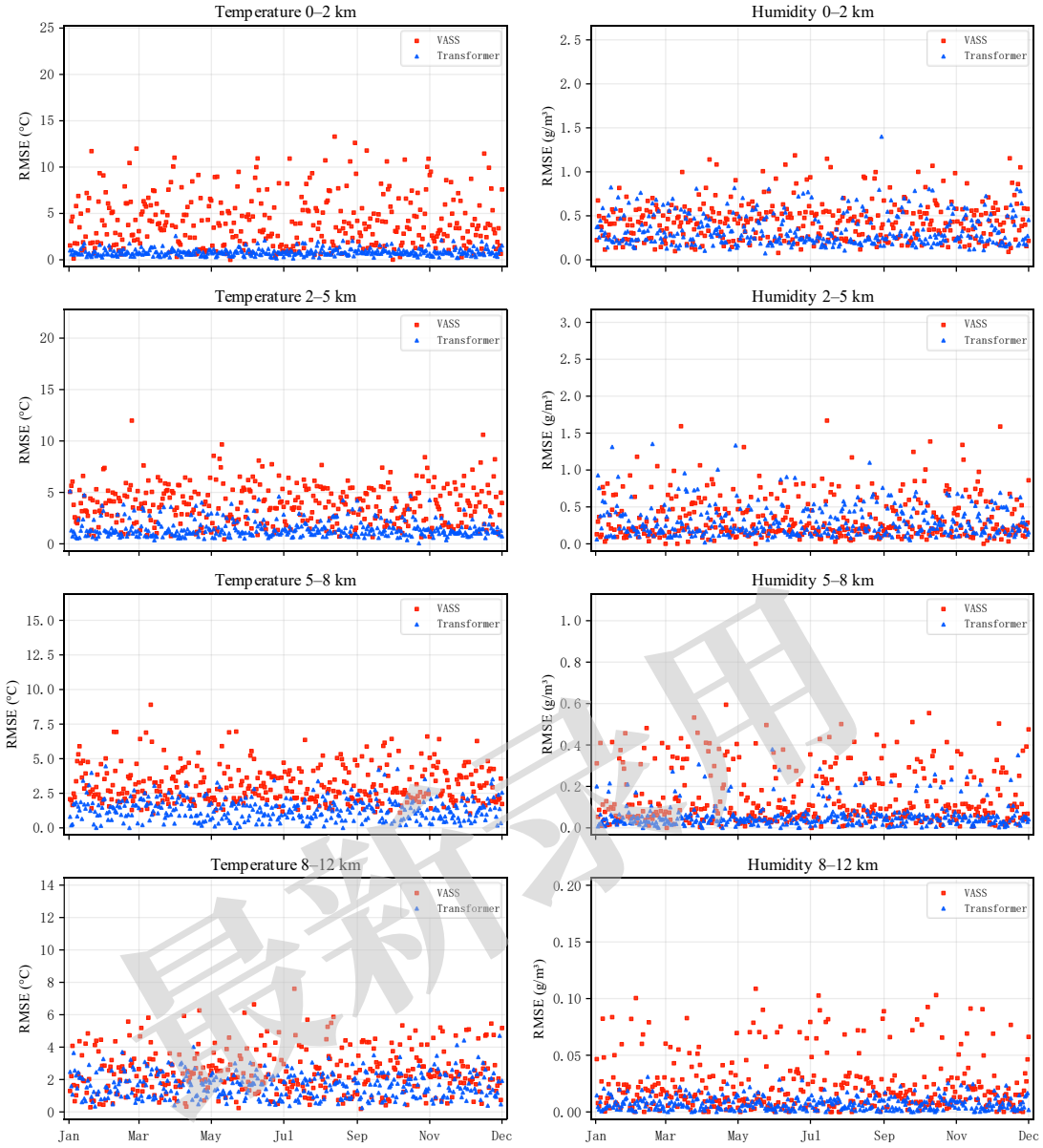


图 10 2025 年 Transformer 反演、FY-3E VASS 产品（降轨）与探空真值之间各高度层的 RMSE 对比（左列：温度，右列：绝对湿度，从上到下依次为：0~2 km、2~5 km、5~8 km、8~12 km）

Fig.10 RMSE comparison among Transformer retrieval, FY-3E VASS product (descending orbit), and radiosonde truth at different altitude levels in 2025 (left column: temperature, right column: absolute humidity, top to bottom: 0~2 km、2~5 km、5~8 km、8~12 km)

图 11 给出了 2025 年 12 月 24 日升轨和降轨两个典型个例中，Transformer 反演、FY-3E VASS 产品与探空真值温湿廓线的对比。结果表明，Transformer 反演廓线整体更接近探空观测，而 VASS 产品在部分高度层存在较明显偏差。升轨个例中，Transformer 温度和湿度 RMSE 分别为 1.76°C 和 0.235g/m^3，VASS 产品为 4.307°C 和 0.98g/m^3；降轨个例中，Transformer 温度和湿度 RMSE 分别为 1.39°C 和 0.163g/m^3，VASS 产品为 7.62°C 和 0.409g/m^3。该结果进一步说明，在本文实验条件下 Transformer 模型能够较好地重建温湿廓线垂直结构。

最新公用

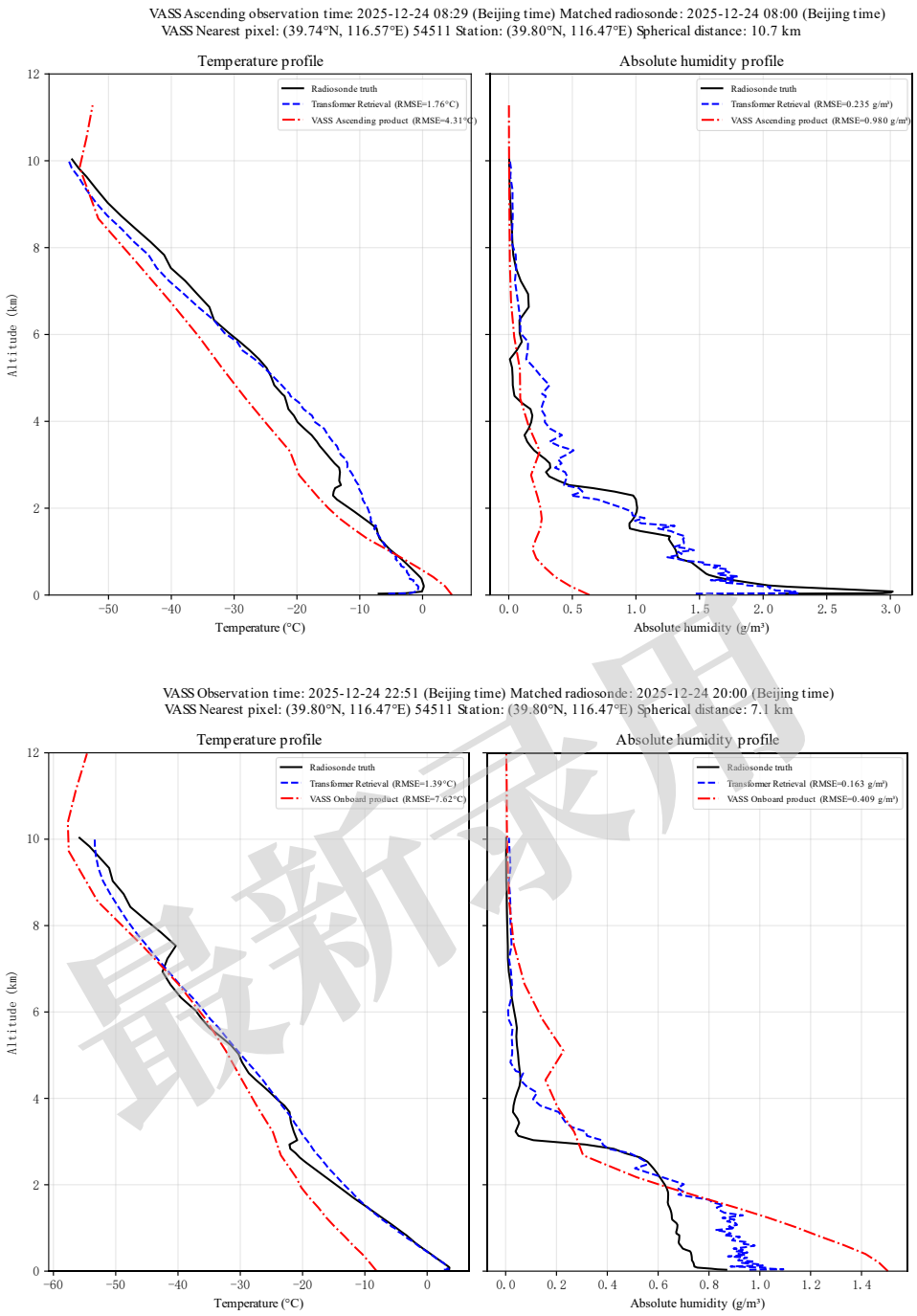


图 11 2025 年 12 月 24 日升轨（上，08:29，匹配 08:00 探空，像元距 10.7km）与降轨（下，22:51，匹配 20:00 探空，像元距 7.1km）温度（左）和绝对湿度（右）廓线对比

Fig.11 Comparison of temperature (left) and absolute humidity (right) profiles for ascending (top, 08:29, matched 08:00 radiosonde, pixel distance 10.7 km) and descending (bottom, 22:51, matched 20:00 radiosonde, pixel distance 7.1 km) orbits on December 24, 2025

4. 结论与展望

基于北京南郊观象台 2020—2025 年探空资料和 MonoRTM 辐射传输模型构建 46 通道模拟亮温数据集，设计了 Sounding Transformer 模型，并与岭回归、BP 神经网络进行了对比。在此基础上，进一步开展了噪声敏感性实验、扩展消融实验、云液态水含量敏感性分析及 FY-3E VASS 产品对比实验，综合评估了模型性能。主要结论如下：

最新录用

（1）将 46 个亮温通道视为具有物理含义的序列输入，利用自注意力机制学习 K/V 波段通道间的耦合关系，并引入时间周期、PWV 和 LWP 等辅助特征，实现了 93 层温度和绝对湿度廓线的联合反演。与岭回归相比，温度 RMSE 降低 22.4%，绝对湿度 RMSE 降低 25.7%；与 BP 神经网络相比，分别降低 3.47%和 2.07%。Sounding Transformer 温度反演优于 BP 神经网络，原因在于温度廓线垂直变化平滑，V 波段亮温与温度的映射规律性强，自注意力机制能够更好地利用通道关联信息；湿度廓线变化剧烈，受云水和边界层过程影响较大，BP 网络已能提取大部分有效信息，Sounding Transformer 的额外增益有限。

（2）在 0~1.75K 噪声范围内，Sounding Transformer 始终保持最优，当噪声增强至 2.0K 以上时，各方法误差均明显增大，Sounding Transformer 温度反演仍优于其他方法，但湿度指标与 BP 网络趋于接近。

（3）时间特征对反演约束最大，去除后温度 RMSE 增大 43.8%；PWV 和 LWP 也有明显贡献，去除后温度 RMSE 增大约 12%，湿度 RMSE 增大约 5%。2 层编码器效果最好，层数过少或过多都会变差。K 波段对云液水敏感，V 波段氧气核心区基本不受影响。

（4）与 FY-3E VASS 产品相比，Sounding Transformer 温度与绝对湿度 RMSE 分别为 1.529°C 和 0.359g/m^3，VASS 分别为 4.127°C 和 0.669g/m^3，升轨和降轨均稳定，近地层改善最突出。这说明该方法可作为探空与卫星之间的验证手段，为风云卫星温湿廓线产品提供验证参考。

本文全部亮温数据均为基于探空资料的 MonoRTM 正演结果，尚未使用真实地基辐射计观测数据进行验证。当 Sounding Transformer 反演方法向真实地基辐射计观测数据迁移时，可能面临以下挑战：（1）辐射计定标误差。通常辐射计定标精度优于 2K，本文在亮温中加入了 0~3K 的高斯噪声进行敏感性分析，该方法适用于定标精度在 2K 以内的辐射计。（2）天线旁瓣与波束填充效应。解决天线旁瓣问题可在天线外部加金属斗抑制旁瓣，或通过扫描点源测量天线旁瓣与波束填充效应所增加的亮温值，并在后续数据处理中消除这部分误差。（3）降水条件下的观测质量。可配合雨量传感器与天线口面维护，建立数据质量控制流程。上述挑战有待真实观测数据验证后逐步解决。

此外，当前模型尚未经过多站点、多气候区的交叉验证，其空间泛化能力尚不明确。推广至其他站点时，可采用迁移学习或分区训练策略。

本方法可进一步推广至其他类型辐射计，如单波段、多波段及具备不同通道配置的辐射计等，并适用于多站点组网观测数据的协同反演。本方法为提高地基辐射计温湿廓线反演精度提供了新的技术途径，有助于提升数值天气预报准确性与空间环境变化监测能力。

参考文献

[1] CIMINI D, WESTWATER E R, GASIEWSKI A J, *et al*. Temperature and humidity profiling in the Arctic using millimeter-wave radiometry[C]//2008 Microwave Radiometry and Remote Sensing of the Environment. Florence: IEEE, 2008: 1-4. DOI: 10.1109/MICRAD.2008.4579484

[2] LIU Hongyan, WANG Yingchun, WANG Jingli, *et al*. Preliminary analysis of the characteristics of precipitable water vapor measured by the ground-based 12-channel microwave radiometer in Beijing[J]. *Chinese Journal of Atmospheric Sciences*, 2009, 33(2): 388-396. DOI: 10.3878/j.issn.1006-9895.2009.02.16 (刘红燕, 王迎春, 王京丽, 等. 由地基微波辐射计测量得到的北京地区水汽特性的初步分析[J]. 大气科学, 2009, 33(2): 388-396. DOI: 10.3878/j.issn.1006-9895.2009.02.16)

[3] LIU Yaya, MAO Jietai, LIU Jun, *et al*. Research of BP neural network for microwave radiometer remote sensing retrieval of temperature, relative humidity, cloud liquid water profiles[J]. *Plateau Meteorology*, 2010, 29(6): 1514-1523 (刘亚亚, 毛节泰, 刘钧, 等. 地基微波辐射计遥感大气廓线的 BP 神经网络反演方法研究[J]. 高原气象, 2010, 29(6): 1514-1523)

[4] WANG D L, TONG L, GONG X, *et al*. Retrieval of atmospheric temperature profiles from hyperspectral microwave radiative data based on the neural network[C]//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Brussels, Belgium: IEEE, 2021: 7095-7098. DOI: 10.1109/IGARSS47720.2021.9554282

[5] YAN X, LIANG C, JIANG Y Z, *et al*. A deep learning approach to improve the retrieval of temperature and humidity profiles from a ground-based microwave radiometer[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(12): 8427-8437. DOI: 10.1109/TGRS.2020.2987896

[6] CHEN Ke, ZHENG Zhaoming, LI Qingxia, *et al*. Three-dimensional atmospheric temperature profile inversion method and system based on DenseNet convolutional neural network: CN, 110826693A[P]. 2020-02-21 (陈柯, 郑照明, 李青侠, 等. 基于 DenseNet 卷积神经网络的三维大气温度廓线反演方法和系统: 中国, 110826693A[P]. 2020-02-21)

[7] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 6000-6010

[8] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An image is worth 16x16 words: transformers for image recognition at scale[C]//9th International Conference on Learning Representations. OpenReview.net, 2021: 1-21
[9] HE K M, CHEN X L, XIE S N, *et al*. Masked autoencoders are scalable vision learners[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 15979-15988. DOI: 10.1109/CVPR52688.2022.01553
[10] Open-Meteo. Historical weather API[DB/OL]. [2026-03-17]. https://open-meteo.com/en/docs/historical-weather-api
[11] World Meteorological Organization. Guide to WMO Table Driven Code Forms (FM 94 BUFR and FM 95 CREX)[R]. Geneva: WMO, 2002
[12] GORDON I E, ROTHMAN L S, HARGREAVES R J, *et al*. The HITRAN2024 molecular spectroscopic database[J]. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2026, 353: 109807
[13] CLOUGH S A, SHEPHARD M W, MLAWER E J, et al. Atmospheric radiative transfer modeling: a summary of the AER codes[J]. Journal of Quantitative Spectroscopy and Radiative Transfer, 2005, 91(2): 233-244. DOI: 10.1016/j.jqsrt.2004.05.058
[14] DEVLIN J, CHANG M W, LEE K, *et al*. BERT: pre-training of deep bidirectional transformers for language understanding[C]//2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: ACL, 2019: 4171-4186
[15] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, *et al*. Dropout: a simple way to prevent neural networks from overfitting[J]. *The Journal of Machine Learning Research*, 2014, 15(1): 1929-1958
[16] XIONG R B, YANG Y C, HE D, et al. On layer normalization in the transformer architecture[C]//Proceedings of the 37th International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2020: 10524-10533
[17] CARUANA R. Multitask learning[J]. *Machine Learning*, 1997, 28(1): 41-75
[18] PEARSON K. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia[J]. Philosophical Transactions of the Royal Society of London A, 1896, 187: 253-318. DOI: 10.1098/rsta.1896.0007
[19] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010: 249-256
[20] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[C]//7th International Conference on Learning Representations. New Orleans: OpenReview.net, 2019: 1-19
[21] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[C]//5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017: 1-16
[22] PRECHELT L. Early stopping — but when?[M]//MONTAVON G, ORR G B, MÜLLER K R. Neural Networks: Tricks of the Trade. 2nd ed. Berlin, Heidelberg: Springer, 2012: 53-67
[23] HORN R A, JOHNSON C R. Matrix Analysis[M]. 2nd ed. Cambridge: Cambridge University Press, 2012: 313-340
[24] ZHANG Xuehui, GUAN Li, WANG Zhenhui, *et al*. Retrieving atmospheric temperature profiles using artificial neural network approach[J]. *Meteorological Monthly*, 2009, 35(11): 137-142 (张雪慧, 官莉, 王振会, 等. 利用人工神经网络方法反演大气温度廓线[J]. 气象, 2009, 35(11): 137-142)
[25] PASZKE A, GROSS S, MASSA F, *et al*. PyTorch: an imperative style, high-performance deep learning library[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: 721

最新录用