



# 基于机器学习的行星大气光谱反演

Planetary Atmospheric Spectrum Retrieval based on Machine Learning

王艺霖

2024年12月26日

# 目录

1 / 背景：行星大气发射光谱

2 / 实现过程

3 / 结果&讨论

# 考虑多次散射的辐射传输方程

$$-\mu \frac{dI(\mu, \phi)}{d\tau^*} = -I + (1 - a)B + a \int \frac{P(\cos \Theta)}{4\pi} I(\mu', \phi') d\Omega'$$

- 因子  $P(\cos \Theta)/4\pi$  表示  $(\mu', \phi')$  方向入射的辐射亮度散射到  $(\mu, \phi)$  方向上的比重

- 应满足  $\int \frac{P(\cos \Theta)}{4\pi} d\Omega = \int \frac{P(\cos \Theta)}{4\pi} d\Omega' = 1$

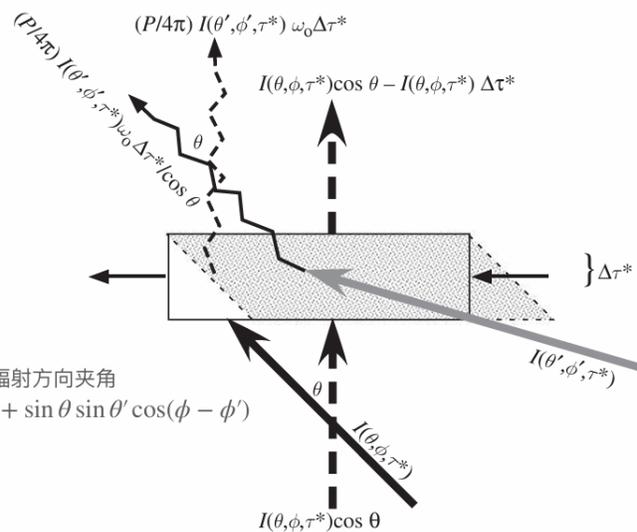
- 各向同性散射  $P(\cos \Theta) \equiv 1$

- Rayleigh phase function

$$P(\cos \Theta) = \frac{3}{2(2 + \delta)} (1 + \delta + (1 - \delta)\cos^2 \Theta)$$

$\Theta$  为入射辐射与散射辐射方向夹角  
 $\cos \Theta = \vec{n}_I \cdot \vec{n}_{sca} = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi')$

10

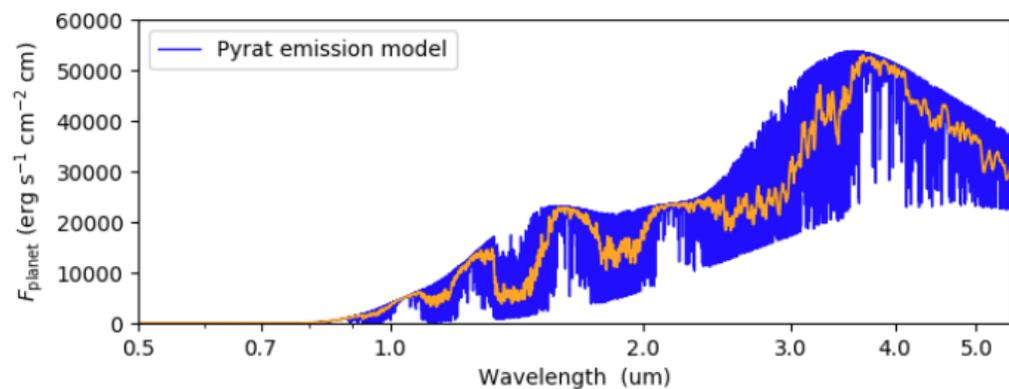


LBL, correlated-k, etc.

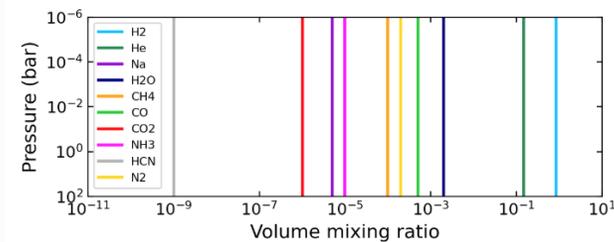
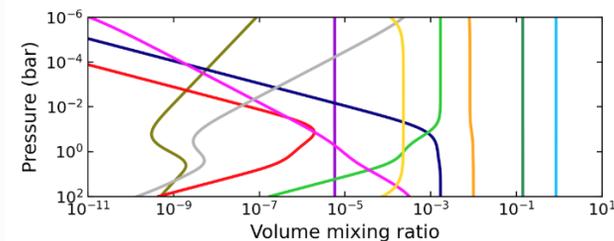
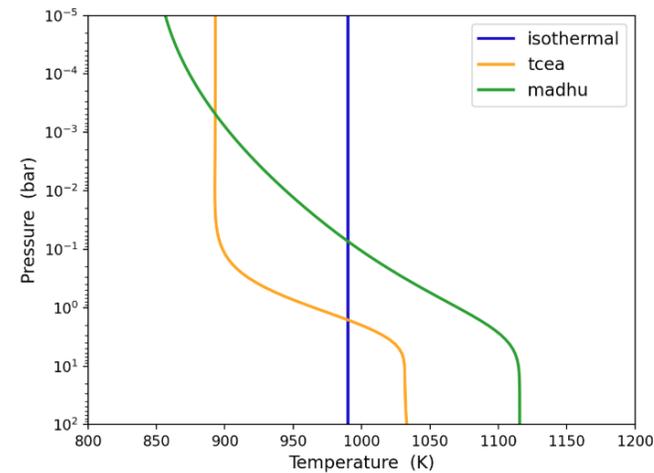
计算量很大，但是没有原则上的困难

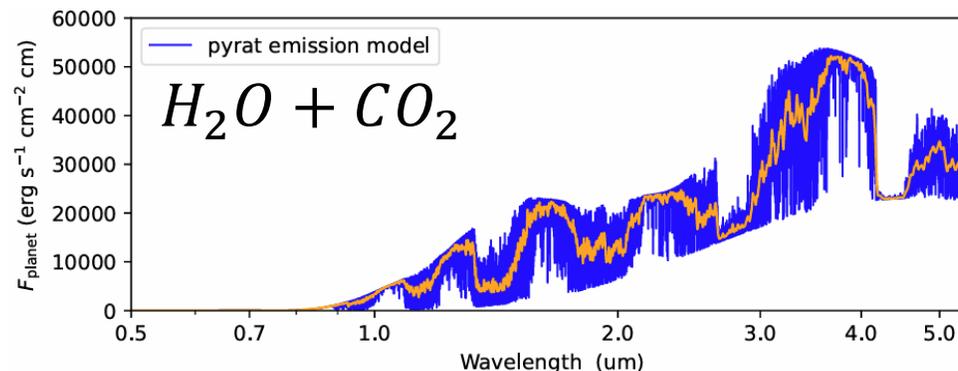
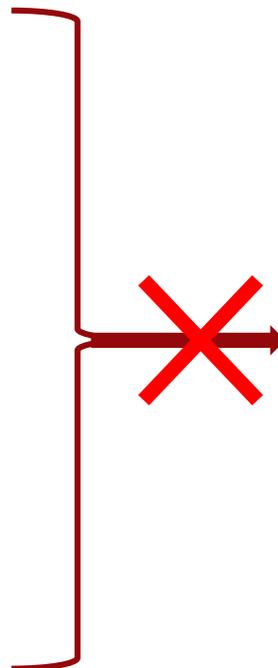
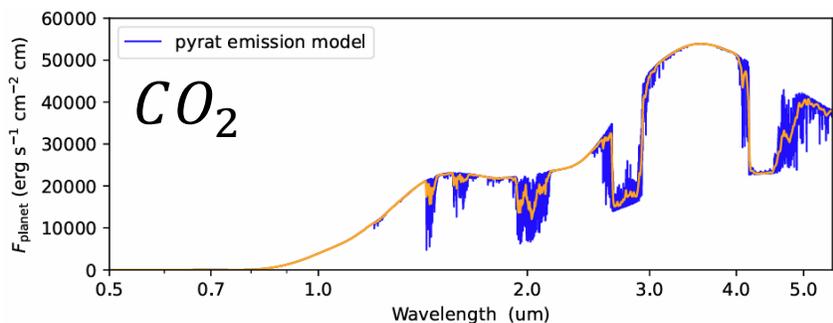
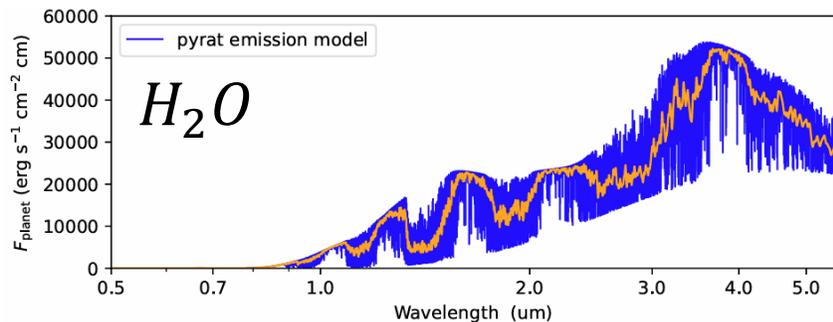
图片来源：丁峰老师

$$p, T(p), X_i(p) \rightarrow F_p(\lambda)$$



$$F_p(\lambda) \rightarrow p, T(p), X_i(p)$$





$$F_p(\lambda) \rightarrow p, T(p), X_i(p)$$

$F_p(\lambda)$  与  $p, T(p), X_i(p)$  的关系是高度非线性的

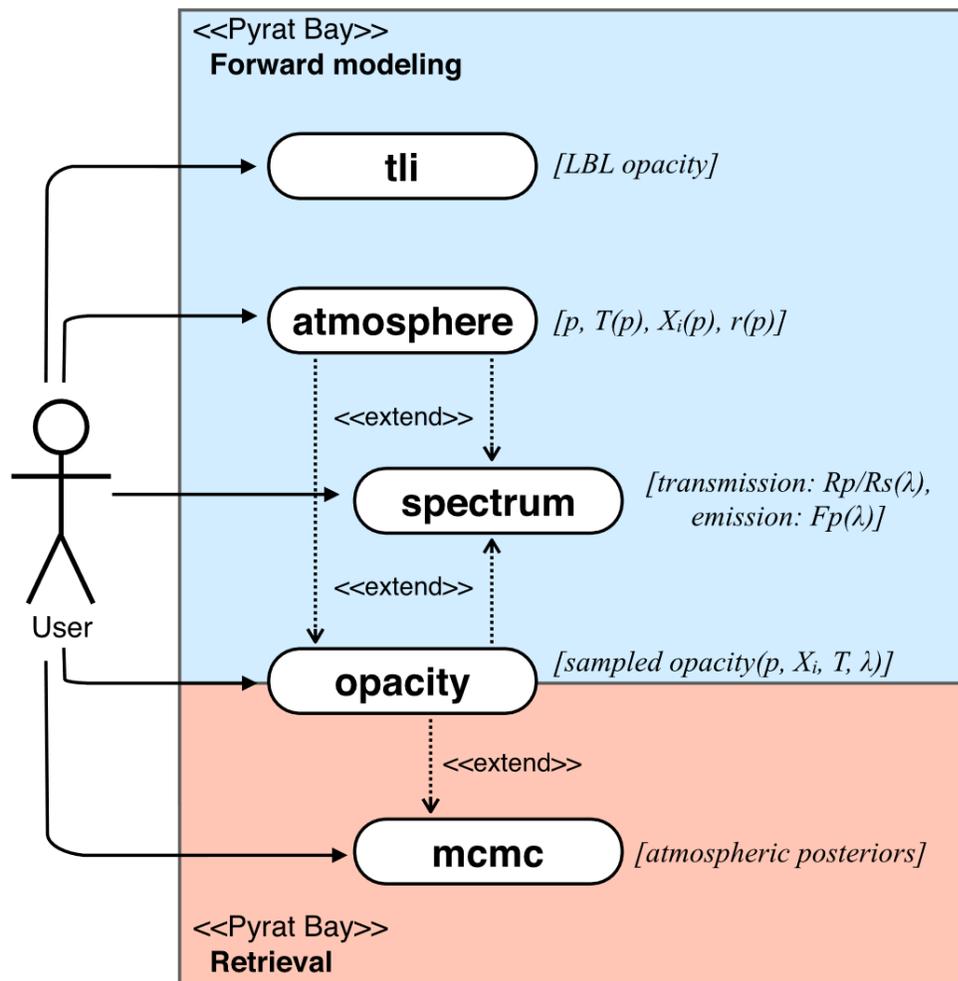
主流方法: **MCMC** (Markov Chain Monte-Carlo) (缺点: 慢, 每一步随机游走都需要计算一次光谱)

## 目录

1 / 背景

2 / 实现过程：数据生成&训练

3 / 结果&讨论



Cubillos & Bleic 2021

## 1. 配置环境

## 2. 从HITRAN下载LBL、CIA的吸收数据

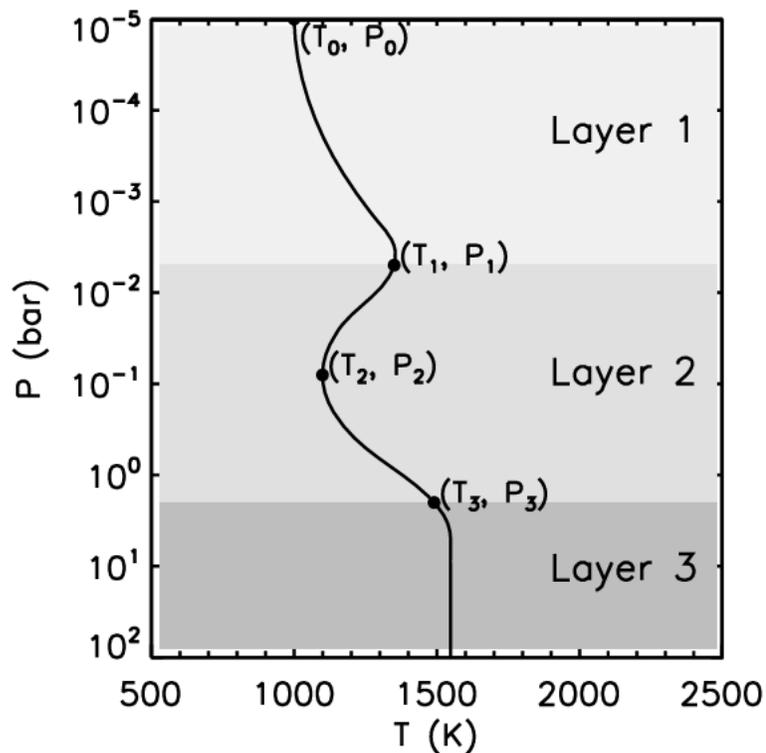
## 3. 单次运行 (略)

- `pbay -c *.cfg`
- `atmos.cfg → profile.atm`
- `spec.cfg → spec.dat`

## 4. 批处理 (略)

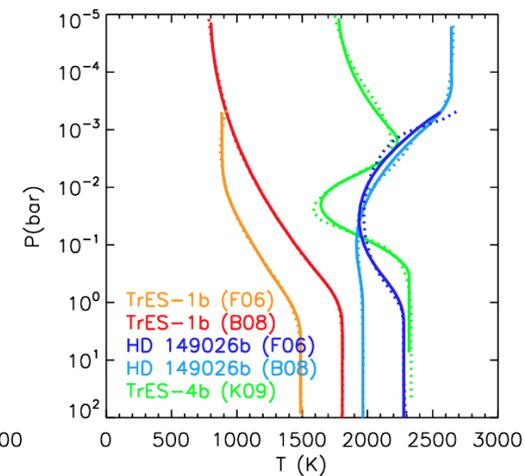
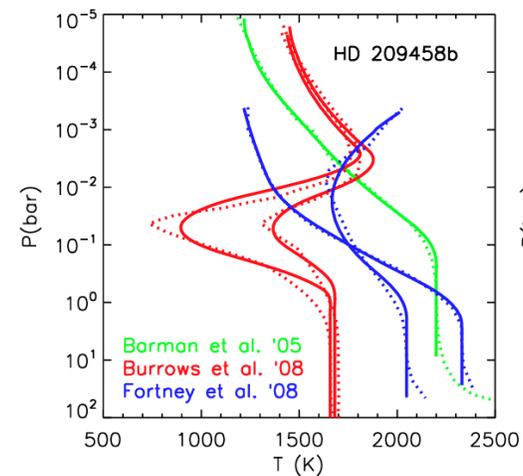
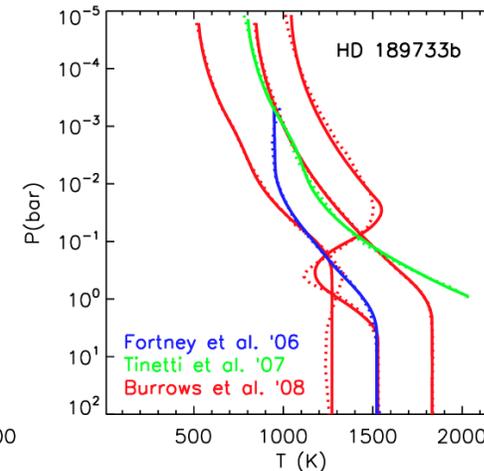
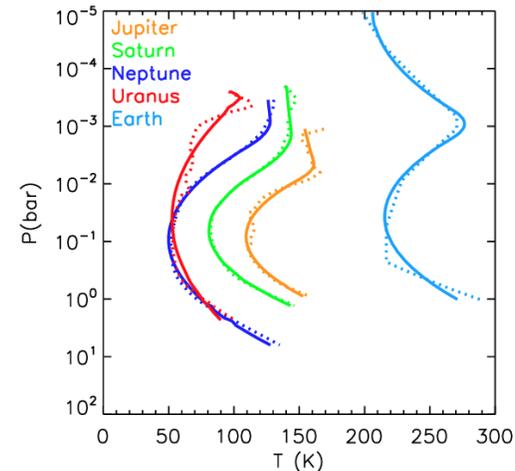
- 生成参数文件 (随机生成)
- 总任务脚本传递task\_id给单个任务脚本
- 单个任务脚本给py程序传递task\_id生成cfg, 并执行pbay命令生成光谱

# 温度廓线: Madhu profile



$$T(p) = \begin{cases} T_0 + \left[ \frac{1}{a_1} \ln(p/p_0) \right]^2 & \text{if } p < p_1 & \text{(layer 1)} \\ T_2 + \left[ \frac{1}{a_2} \ln(p/p_2) \right]^2 & \text{if } p_1 \leq p < p_3 & \text{(layer 2)} \\ T_3 & \text{if } p \geq p_3 & \text{(layer 3)} \end{cases}$$

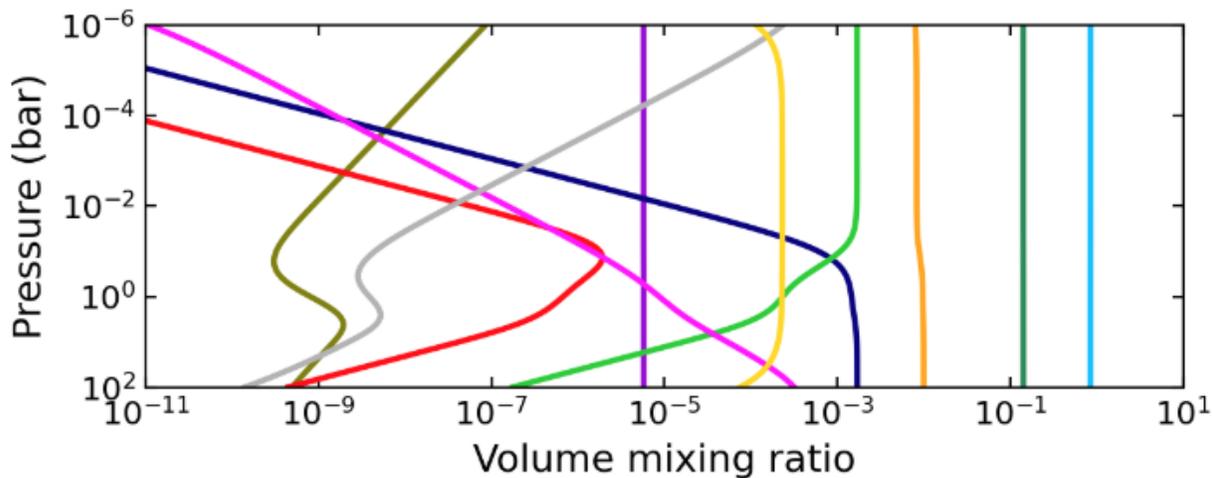
若  $p_1 < p_2$ , 则有逆温层, 反之则无



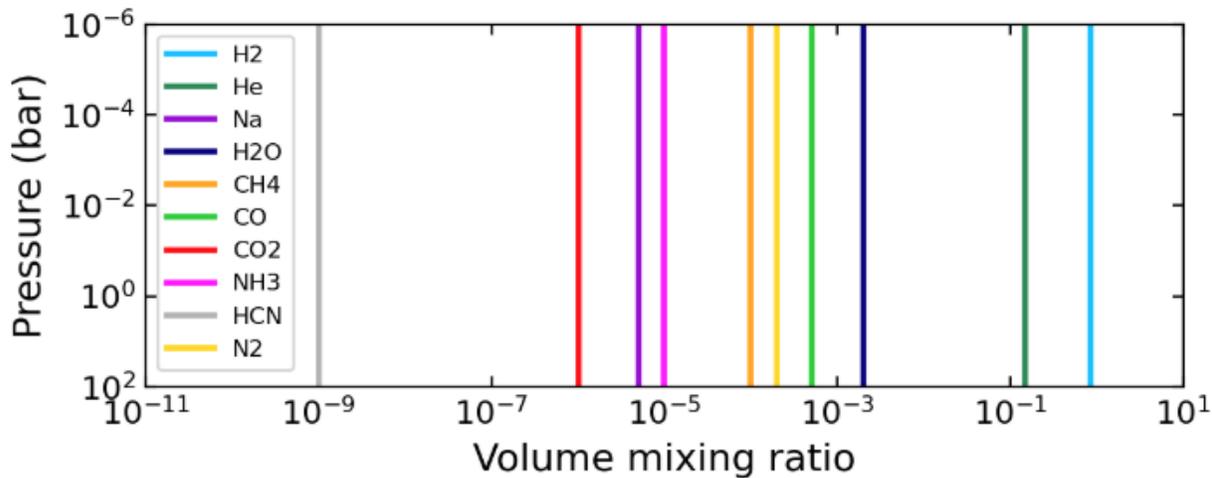
Madhusudhan & Seager 2009

这个模型是为了拟合系外行星温度廓线提出的

# 大气成分浓度廓线：均匀分布



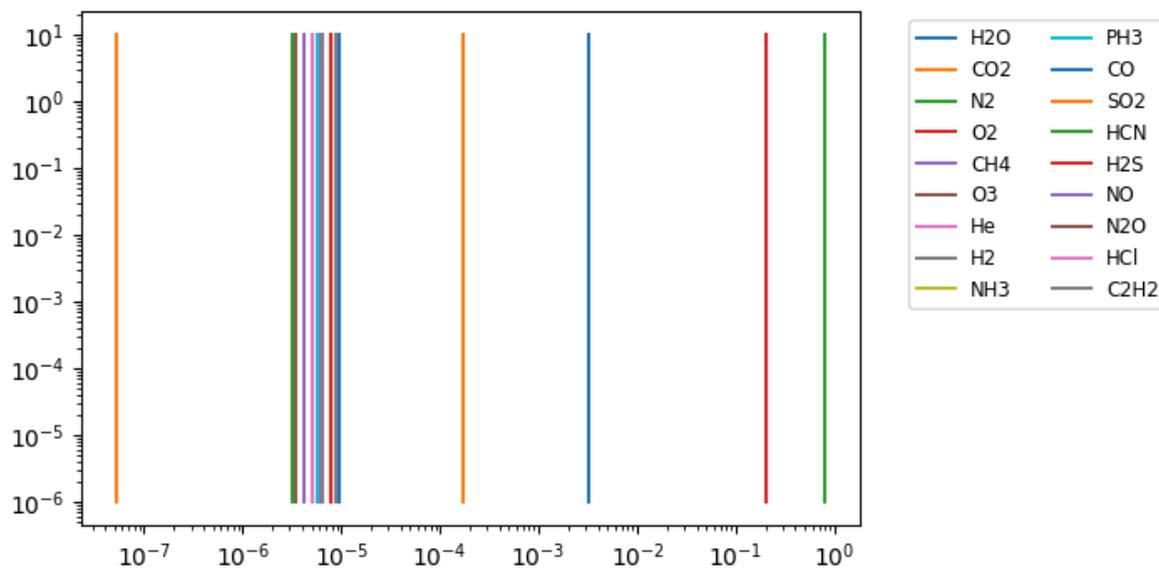
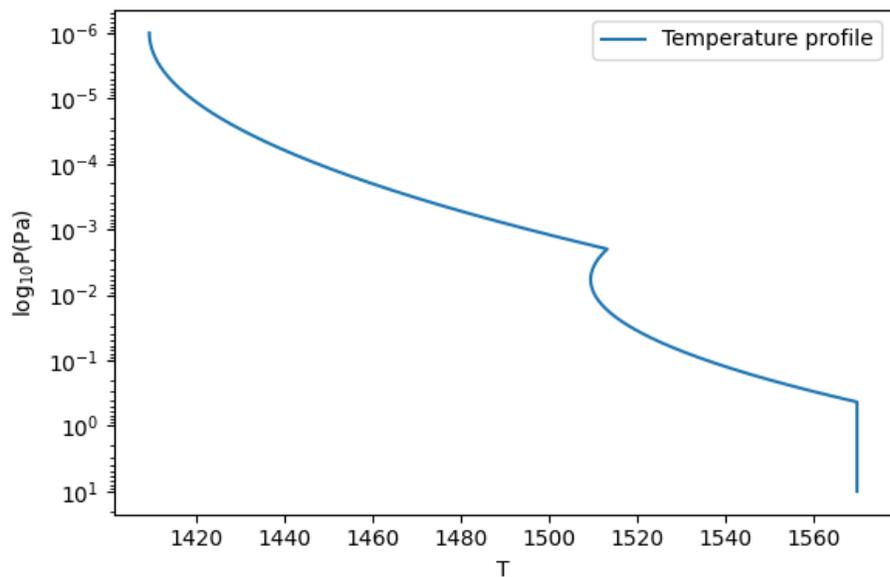
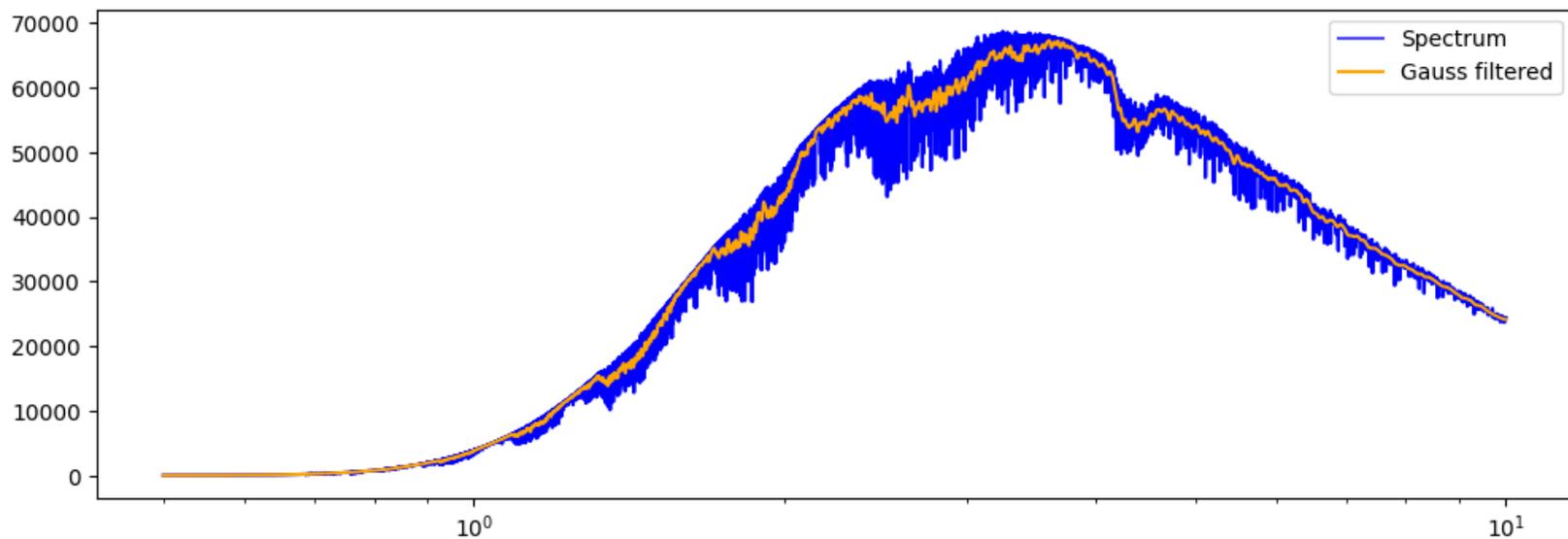
Thermal Equilibrium



**Uniform**

为简单起见，选择均匀分布

# 一个数据点实例



- Madhu profile的6个参数（后来证明这是个错误的选择）

$$p_1, p_2, p_3, a_1, a_2, T_0$$

- 18种气体的含量（8种主要气体+10种微量气体）

$H_2O, CO_2, N_2, O_2, CH_4, O_3, He, H_2, NH_3, PH_3, CO, SO_2, HCN, H_2S, NO, N_2O, HCl, C_2H_2$

- 总共生成10800组数据
- 输入数据：X (10800, 19001)
- 预测参量：Y (10800, 24)

尝试6种模型：

**线性回归、随机森林、MLP、1D CNN、2D CNN、Transformer**

均采用 train:validation:test=8:1:1 随机划分数据

- 2D CNN:

把 $x_i$ (19001)折叠为 $(190 \times 100)$ 的二维数组

优点：可以使用二维卷积核，更好地利用了CNN的特性

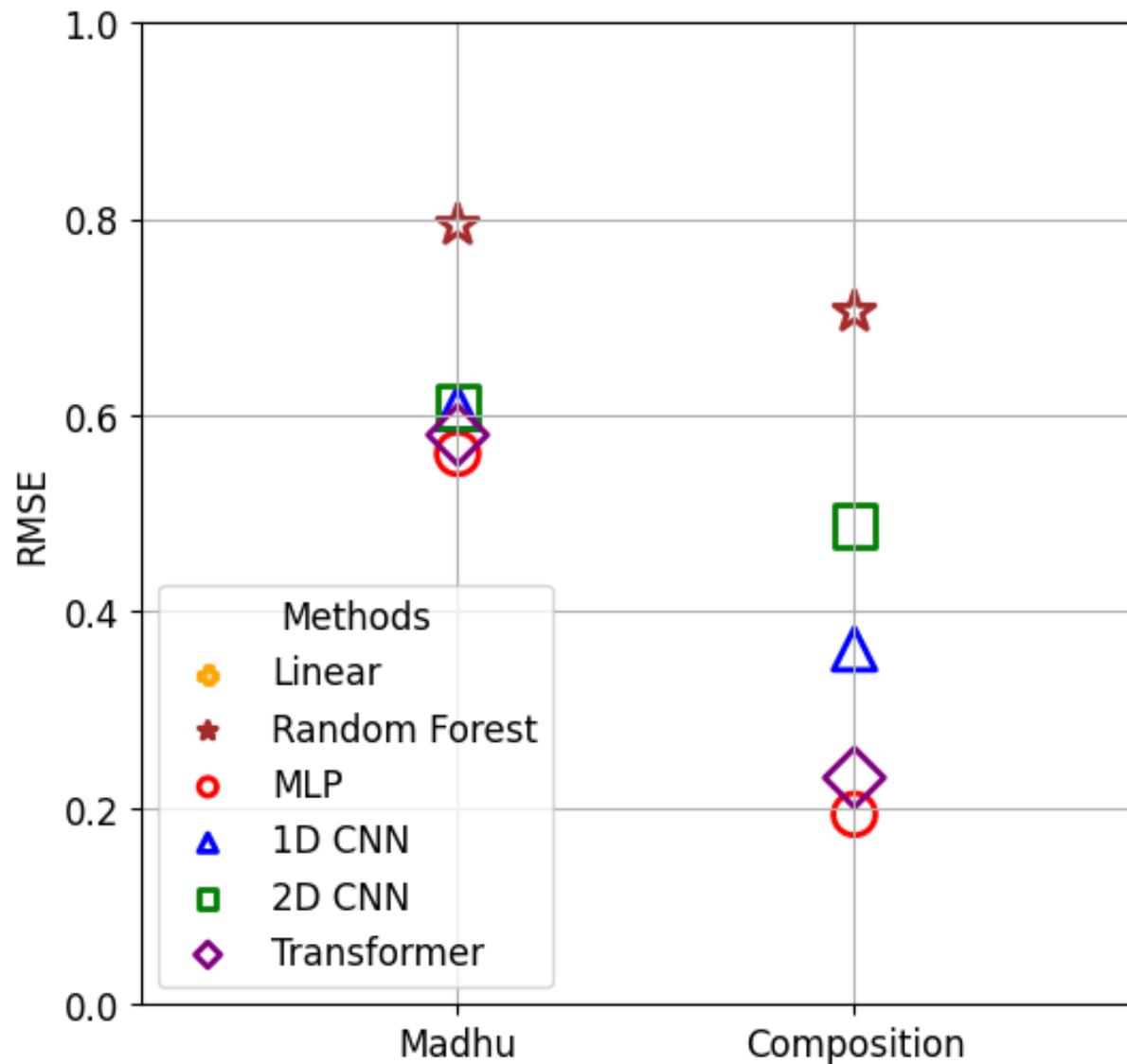
缺点：丢失了原本连接处的序列信息

# 目录

1 / 背景

2 / 实现过程

3 / 结果&讨论

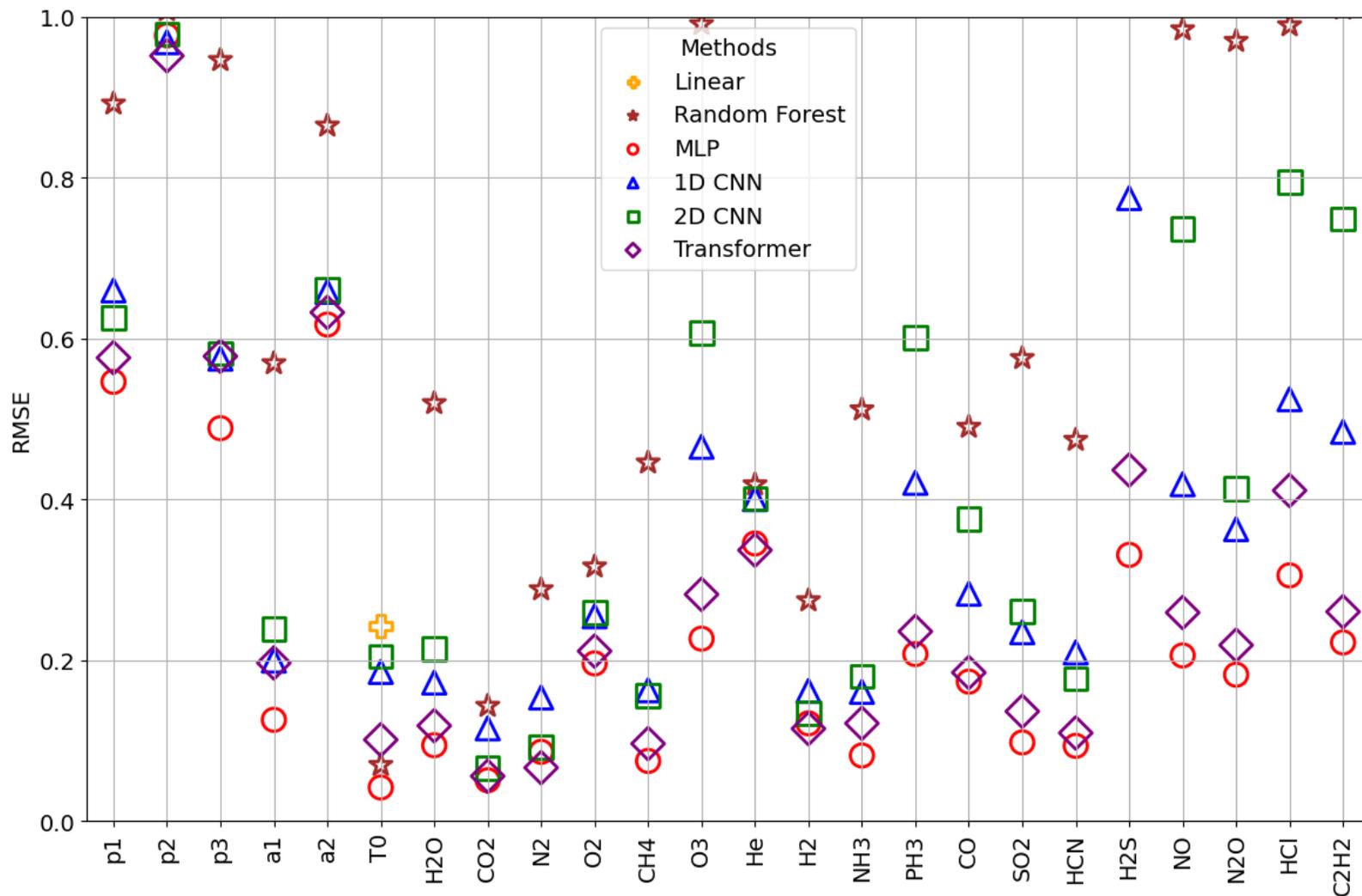


幽默的线性回归：

- 前 6 个参数的 RMSE: 76.13
- 后 18 个参数的 RMSE: 43.74

(也反映了光谱和大气的关系的确是高度非线性的)

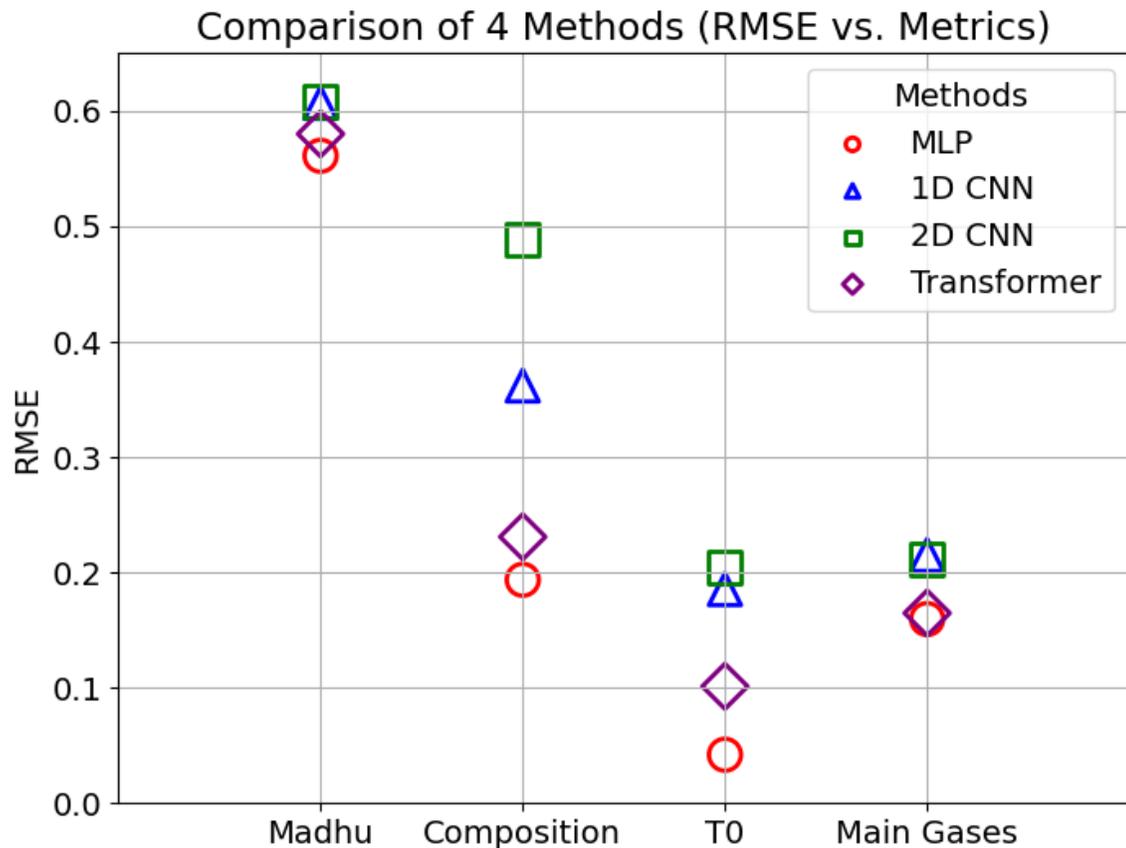
# 重要参数： $T_0$ & 主要气体



发现 $a_1, T_0$ 和8种主要气体的含量的预测效果还可以，其余参量的预测效果较差

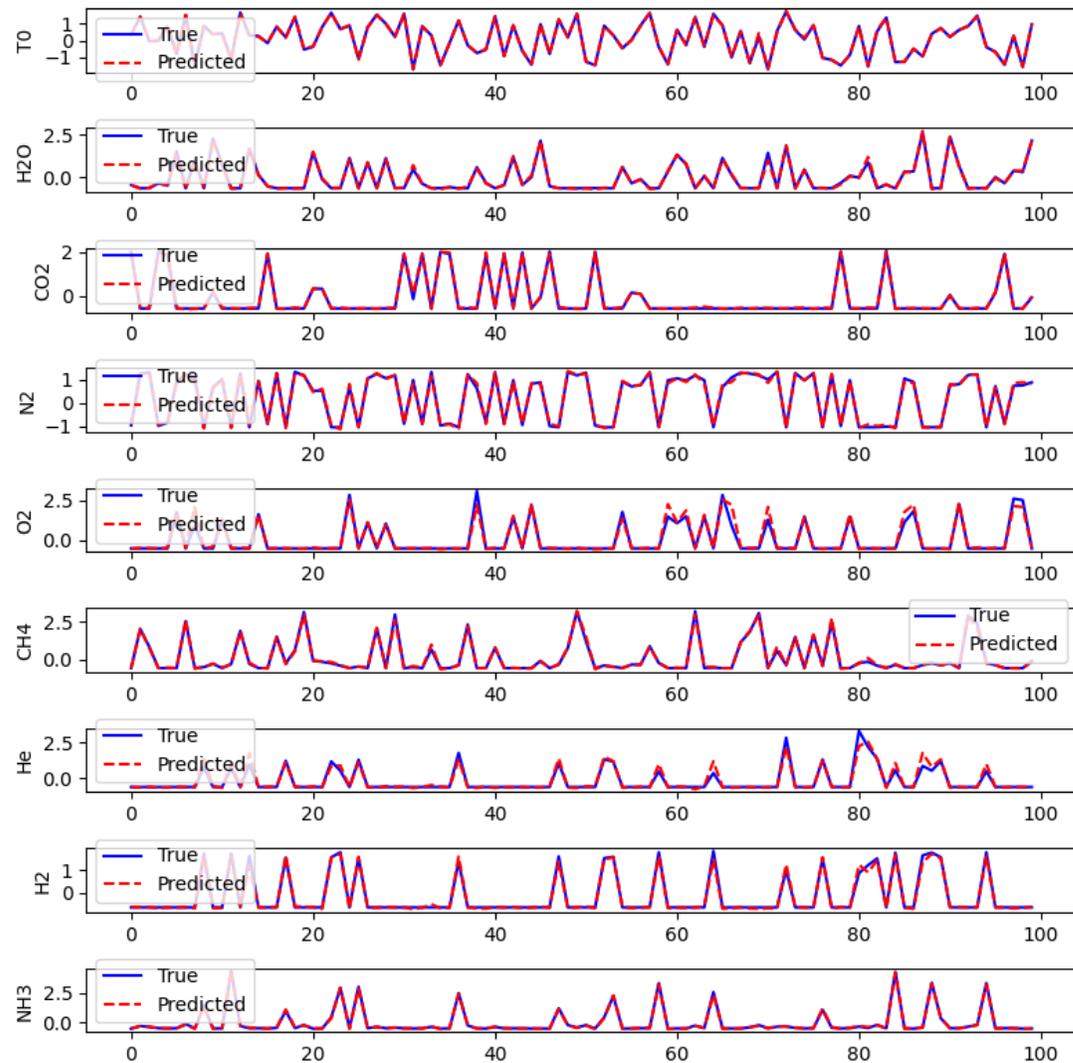
这是可以理解的，对光谱影响小的参数很难在 $10^4$ 个样本中学习

重要参数： $T_0, \{H_2O, CO_2, N_2, O_2, CH_4, He, H_2, NH_3\}$

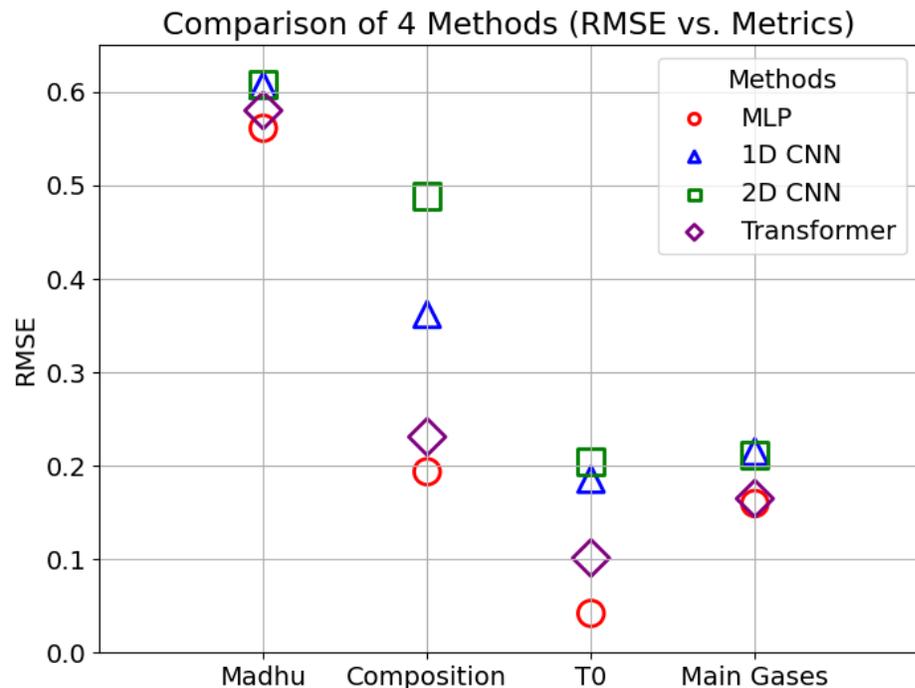


效果还不错！（考虑到只有 $10^4$ 的样本量和极为有限的调参次数）

（以及，把一维数据折叠成二维来套用2D CNN看起来是损失大于收益的）



MLP预测结果（100个测试点）



- **温度廓线**: Madhu参数对于温度廓线的影响比较复杂; 更重要的影响因素是**光学厚度**, 在发射光谱中光学厚度为1位置的权重函数最大, 深层大气发射的光谱无法穿透 (Madhu参数中的 $a_1$ 对这个位置的决定也有作用, 所以学的效果明显好于剩余的4个参数, 剩下4个参数主要对更深的大气有影响)
- **大气成分**: 大概能判断出来行星大气的主要成分和含量, 但肯定与biosignature识别这种任务无缘

接下来的改进方向:

- 使用每一层的温度而不是Madhu参数作为温度廓线的输入变量
- 使用达到thermal equilibrium的气体成分分布
- 和主流的方法的预测效果、资源消耗进行对比

1. Cubillos, P. E., & Blecic, J. (2021). The pyrat bay framework for exoplanet atmospheric modelling: A population study of Hubble/WFC3 transmission spectra. *Monthly Notices of the Royal Astronomical Society*, 505(2), 2675–2702. <https://doi.org/10.1093/mnras/stab1405>
2. Madhusudhan, N., & Seager, S. (2009). A temperature and abundance retrieval method for exoplanet atmospheres. *The Astrophysical Journal*, 707(1), 24. <https://doi.org/10.1088/0004-637X/707/1/24>
3. Line, M. R., Wolf, A. S., Zhang, X., Knutson, H., Kammer, J. A., Ellison, E., Deroo, P., Crisp, D., & Yung, Y. L. (2013). A systematic retrieval analysis of secondary eclipse spectra. I. A comparison of atmospheric retrieval techniques. *The Astrophysical Journal*, 775(2), 137. <https://doi.org/10.1088/0004-637X/775/2/137>

**感谢大家！**

欢迎访问 <https://williamwyl.cn> ! (slides+data+source code)