

# Supplementary Materials

## Pressure-Level-Guided Spectral Cross-Attention for AIRS Temperature Profile Retrieval over East Asia under Clear-Sky Summer Conditions

Yujian Zhou et al.

### Supplementary Methods S1. Detailed PC-CAN Architecture

#### S1.1 Residual spectral encoder

The input-tensor notation in Equation S1 is defined in this study. For a batch of AIRS radiance vectors, a singleton channel dimension is introduced so that

$$\mathbf{X}_0 \in \mathbb{R}^{B \times 1 \times 2645}. \quad (\text{S1})$$

The stem and residual mappings in Equations S2–S4 adapt the ResNet formulation of He et al. [1] to one-dimensional AIRS spectral inputs. The stem applies a kernel-7, stride-1 convolution, batch normalization, and ReLU:

$$\mathbf{F}_0 = \text{ReLU}[\text{BN}(\text{Conv}_7(\mathbf{X}_0))]. \quad (\text{S2})$$

Three stages then apply two BasicBlock1D modules followed by kernel-2, stride-2 max pooling. For input feature  $\mathbf{x}$ , the residual block is

$$\mathbf{y} = \text{ReLU}\left\{\text{BN}_2\left[\text{Conv}_3^{(2)}\left(\text{Dropout1D}\left[\text{ReLU}\left(\text{BN}_1[\text{Conv}_3^{(1)}(\mathbf{x})]\right)\right]\right)\right] + \phi(\mathbf{x})\right\}, \quad (\text{S3})$$

with shortcut

$$\phi(\mathbf{x}) = \begin{cases} \mathbf{x}, & C_{\text{in}} = C_{\text{out}}, \\ \text{BN}[\text{Conv}_1(\mathbf{x})], & C_{\text{in}} \neq C_{\text{out}}. \end{cases} \quad (\text{S4})$$

Table S1: Detailed structure of the one-dimensional residual spectral encoder.

| Module  | Output channels | Residual blocks | Spectral length |
|---------|-----------------|-----------------|-----------------|
| Stem    | 32              | —               | 2645            |
| Stage 1 | 64              | 2               | 1322            |
| Stage 2 | 128             | 2               | 661             |
| Stage 3 | 256             | 2               | 330             |

#### S1.2 Spectral tokenization and pressure queries

The fixed-length spectral-token operator in Equation S5 is defined here as part of PC-CAN. A  $1 \times 1$  convolution reduces the 256 encoder channels to 64, and adaptive average pooling converts the spectral length from 330 to 64:

$$\mathbf{S} = \text{Transpose}\{\text{AdaptiveAvgPool1D}_{64}[\text{Conv}_1(\mathbf{F})]\} \in \mathbb{R}^{B \times 64 \times 64}. \quad (\text{S5})$$

Equations S6 and S7 define the normalized log-pressure coordinate and the empirical pressure basis used in this study. The trigonometric terms in Equation S7 are inspired by sinusoidal positional

encoding [2]; the complete seven-dimensional basis and its use as a pressure-conditioned query are specific to PC-CAN. For pressure level  $p_l$ , the continuous coordinate and seven-dimensional empirical basis are

$$u_l = 2 \frac{\ln p_l - \ln p_{\min}}{\ln p_{\max} - \ln p_{\min} + \varepsilon} - 1, \quad (\text{S6})$$

$$\mathbf{e}_l = [u_l, u_l^2, u_l^3, \sin(\pi u_l), \cos(\pi u_l), \sin(2\pi u_l), \cos(2\pi u_l)]. \quad (\text{S7})$$

A shared linear layer maps  $\mathbf{e}_l$  to a 64-dimensional query. Stacking and broadcasting the 37 queries gives  $\mathbf{Q} \in \mathbb{R}^{B \times 37 \times 64}$ . This basis is a pressure-coordinate conditioning design and is not a constraint derived from radiative transfer.

### S1.3 Cross-attention, feed-forward block, and regression head

Equations S8–S12 follow the multi-head scaled dot-product attention and residual feed-forward framework of Vaswani et al. [2], adapted here so that pressure-conditioned vectors serve as queries and AIRS spectral tokens serve as keys and values. For each of four heads,

$$\mathbf{Q}_h = \mathbf{Q} \mathbf{W}_h^Q, \quad \mathbf{K}_h = \mathbf{S} \mathbf{W}_h^K, \quad \mathbf{V}_h = \mathbf{S} \mathbf{W}_h^V, \quad (\text{S8})$$

$$\mathbf{A}_h = \text{softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^\top}{\sqrt{d_h}}\right), \quad \mathbf{O}_h = \mathbf{A}_h \mathbf{V}_h, \quad (\text{S9})$$

where  $d_h = 16$ . The four outputs are concatenated and projected:

$$\mathbf{O} = \text{Concat}(\mathbf{O}_1, \dots, \mathbf{O}_4) \mathbf{W}^O. \quad (\text{S10})$$

Post-layer normalization and residual connections are applied as

$$\mathbf{Z} = \text{LayerNorm}(\mathbf{Q} + \mathbf{O}), \quad (\text{S11})$$

$$\mathbf{H} = \text{LayerNorm}[\mathbf{Z} + \text{FFN}(\mathbf{Z})], \quad (\text{S12})$$

The feed-forward network expands the feature dimension from 64 to 256, followed by GELU and dropout, and then projects it back to 64 features with a second dropout operation. The resulting tensor  $\mathbf{H} \in \mathbb{R}^{B \times 37 \times 64}$  is processed by the same regression head at every level: The shared level-wise regression head in Equations S13 and S14 is defined in this study to map each pressure-specific feature to one temperature estimate.

$$\mathbf{r}_{i,l} = \text{GELU}(\mathbf{h}_{i,l} \mathbf{W}_1 + \mathbf{b}_1), \quad (\text{S13})$$

$$\hat{y}_{i,l} = \text{Dropout}(\mathbf{r}_{i,l}) \mathbf{W}_2 + \mathbf{b}_2, \quad (\text{S14})$$

with  $\mathbf{W}_1 \in \mathbb{R}^{64 \times 64}$  and  $\mathbf{W}_2 \in \mathbb{R}^{64 \times 1}$ .

## Supplementary Table S2. Candidate Model Configurations

All candidates were evaluated only on the validation set. Because token count and token dimension varied together in part of the search, the first three rows are comparisons among combined configurations rather than a strict single-factor ablation.

Table S2: Validation RMSE for the evaluated PC-CAN configurations.

| Comparison          | Candidate | Tokens | Dimension | Heads | Validation RMSE (K) |
|---------------------|-----------|--------|-----------|-------|---------------------|
| Token configuration | A         | 64     | 64        | 4     | 1.4252              |
| Token configuration | B         | 96     | 128       | 4     | 1.5886              |
| Token configuration | C         | 128    | 128       | 4     | 1.5782              |
| Feature dimension   | A         | 64     | 64        | 4     | 1.4252              |
| Feature dimension   | D         | 64     | 192       | 4     | 1.4891              |
| Attention heads     | E         | 64     | 64        | 2     | 1.6298              |
| Attention heads     | A         | 64     | 64        | 4     | 1.4252              |
| Attention heads     | F         | 64     | 64        | 8     | 1.6198              |

### Supplementary Table S3. Valid Test-Sample Counts

A pressure level is valid only when its pressure does not exceed the ERA5 surface pressure interpolated to the AIRS footprint. Fractions are relative to all 84,618 test profiles.

Table S3: Valid test-sample count and fraction at every standard pressure level.

| Pressure (hPa) | Valid $N$ | Valid (%) | Pressure (hPa) | Valid $N$ | Valid (%) |
|----------------|-----------|-----------|----------------|-----------|-----------|
| 1              | 84,618    | 100.0     | 300            | 84,618    | 100.0     |
| 2              | 84,618    | 100.0     | 350            | 84,618    | 100.0     |
| 3              | 84,618    | 100.0     | 400            | 84,618    | 100.0     |
| 5              | 84,618    | 100.0     | 450            | 84,618    | 100.0     |
| 7              | 84,618    | 100.0     | 500            | 84,618    | 100.0     |
| 10             | 84,618    | 100.0     | 550            | 84,099    | 99.4      |
| 20             | 84,618    | 100.0     | 600            | 81,883    | 96.8      |
| 30             | 84,618    | 100.0     | 650            | 79,945    | 94.5      |
| 50             | 84,618    | 100.0     | 700            | 78,048    | 92.2      |
| 70             | 84,618    | 100.0     | 750            | 75,429    | 89.1      |
| 100            | 84,618    | 100.0     | 775            | 74,202    | 87.7      |
| 125            | 84,618    | 100.0     | 800            | 72,306    | 85.4      |
| 150            | 84,618    | 100.0     | 825            | 69,384    | 82.0      |
| 175            | 84,618    | 100.0     | 850            | 63,287    | 74.8      |
| 200            | 84,618    | 100.0     | 875            | 52,114    | 61.6      |
| 225            | 84,618    | 100.0     | 900            | 41,130    | 48.6      |
| 250            | 84,618    | 100.0     | 925            | 34,528    | 40.8      |
|                |           |           | 950            | 27,439    | 32.4      |
|                |           |           | 975            | 18,220    | 21.5      |
|                |           |           | 1000           | 9,550     | 11.3      |

## References

- [1] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [2] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017; pp. 5998–6008.