npj Artificial Intelligence

Supplementary Information *for*

Efficient Physics-Aware Deep Learning for Skillful Tropical Stratosphere Dynamics Modeling

Ran Tao[1,2,3], Xianghui Xue[1,2,3,*], Hao Chen[4,*], Fenghua Ling[4], Fei Xie[5], Ben Fei[4], Lei Bai[4], and Xiankang Dou[1,2,3]

[1]School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China

[2]School of Emerging Technology, University of Science and Technology of China, Hefei, China

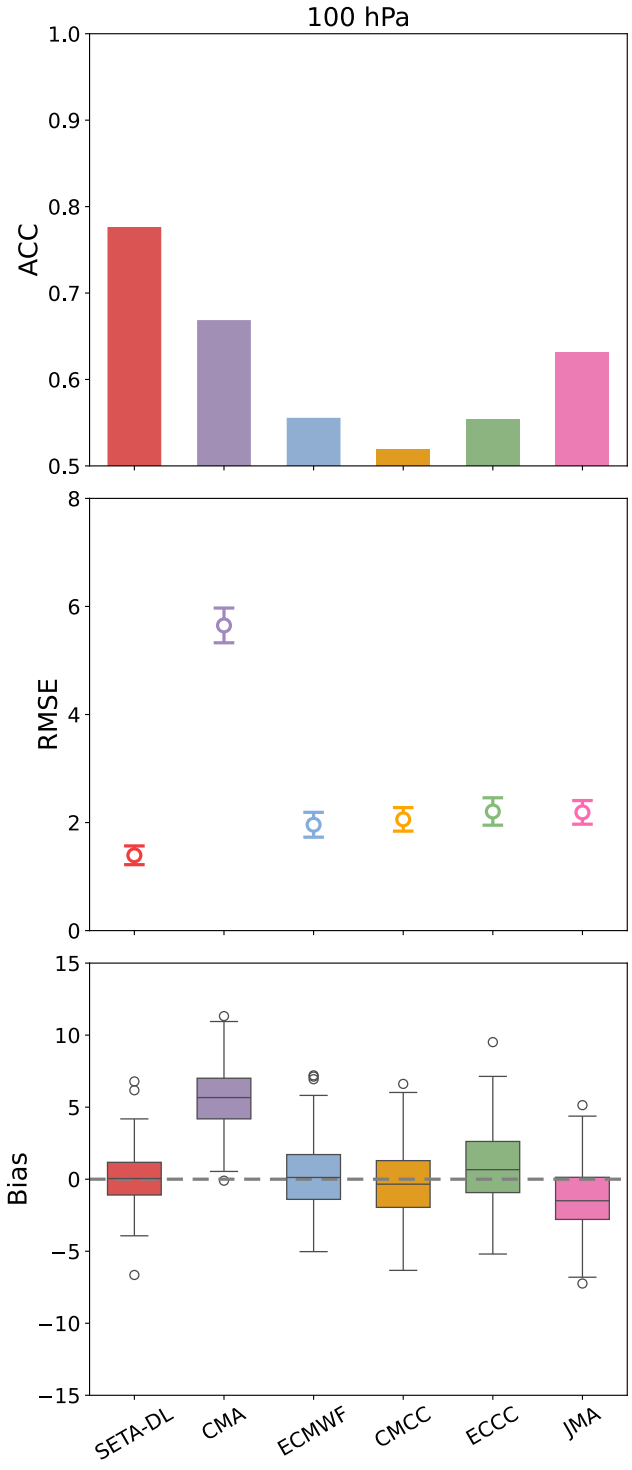
[3]Hefei National Laboratory, Hefei, China

[4]Shanghai Artificial Intelligence Laboratory, Shanghai, China

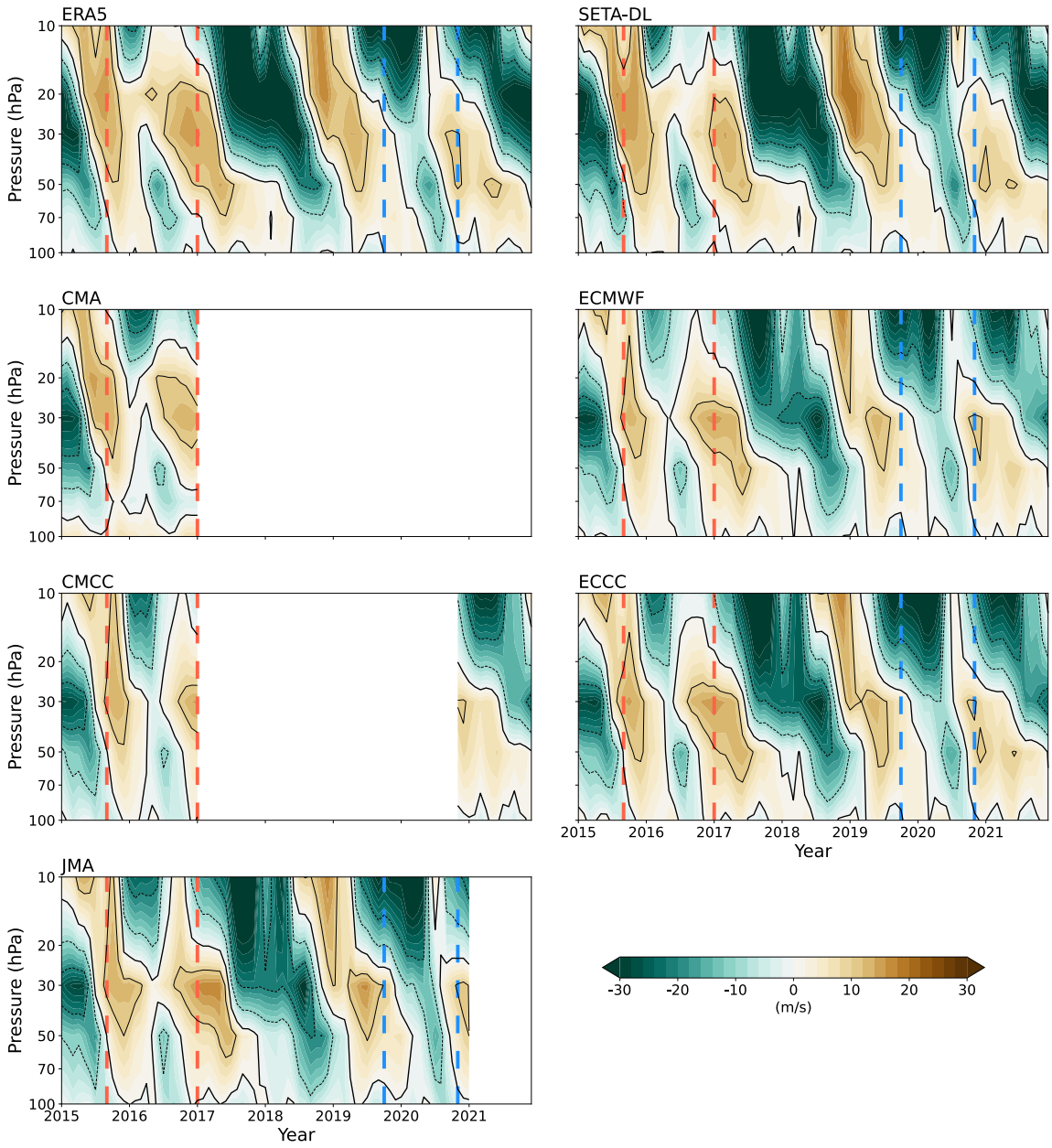
[5]School of Systems Science, Beijing Normal University, Beijing, China

*Corresponding author(s). Email(s): xuexh@ustc.edu.cn; chenhao1@pjlab.org.cn.

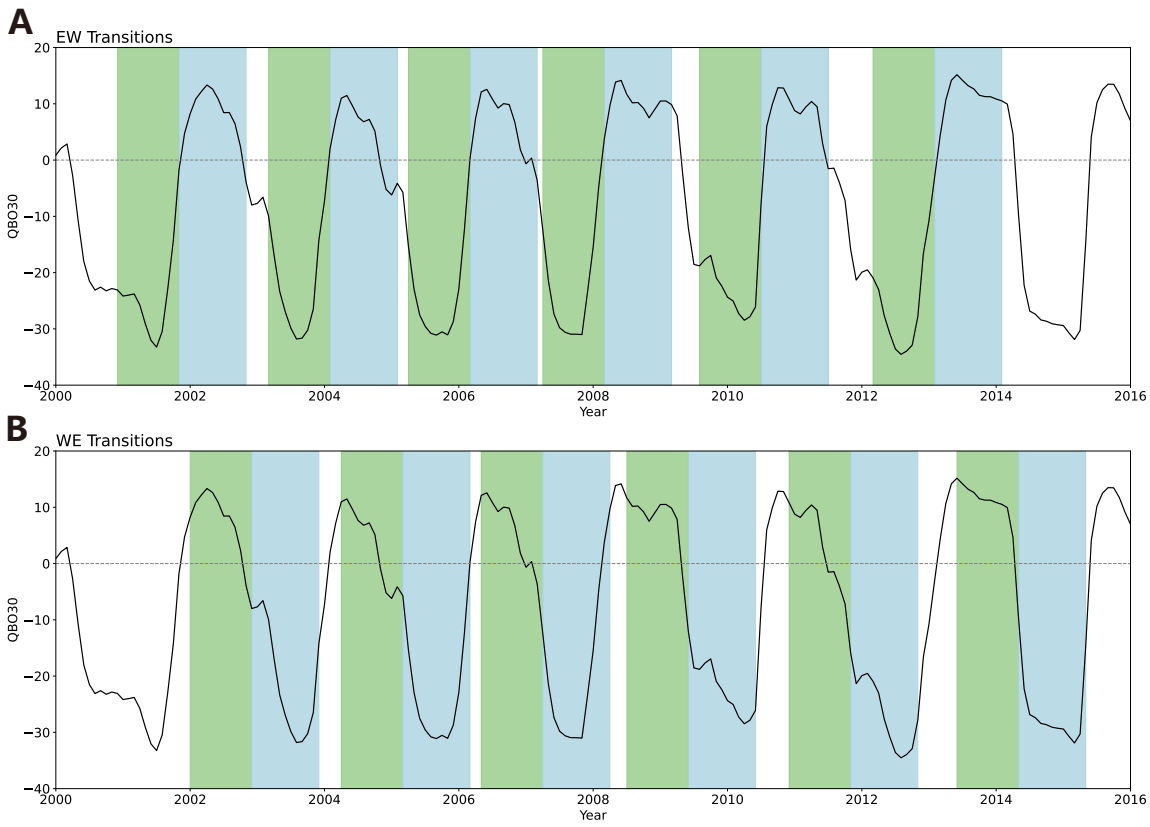
Contents of this file



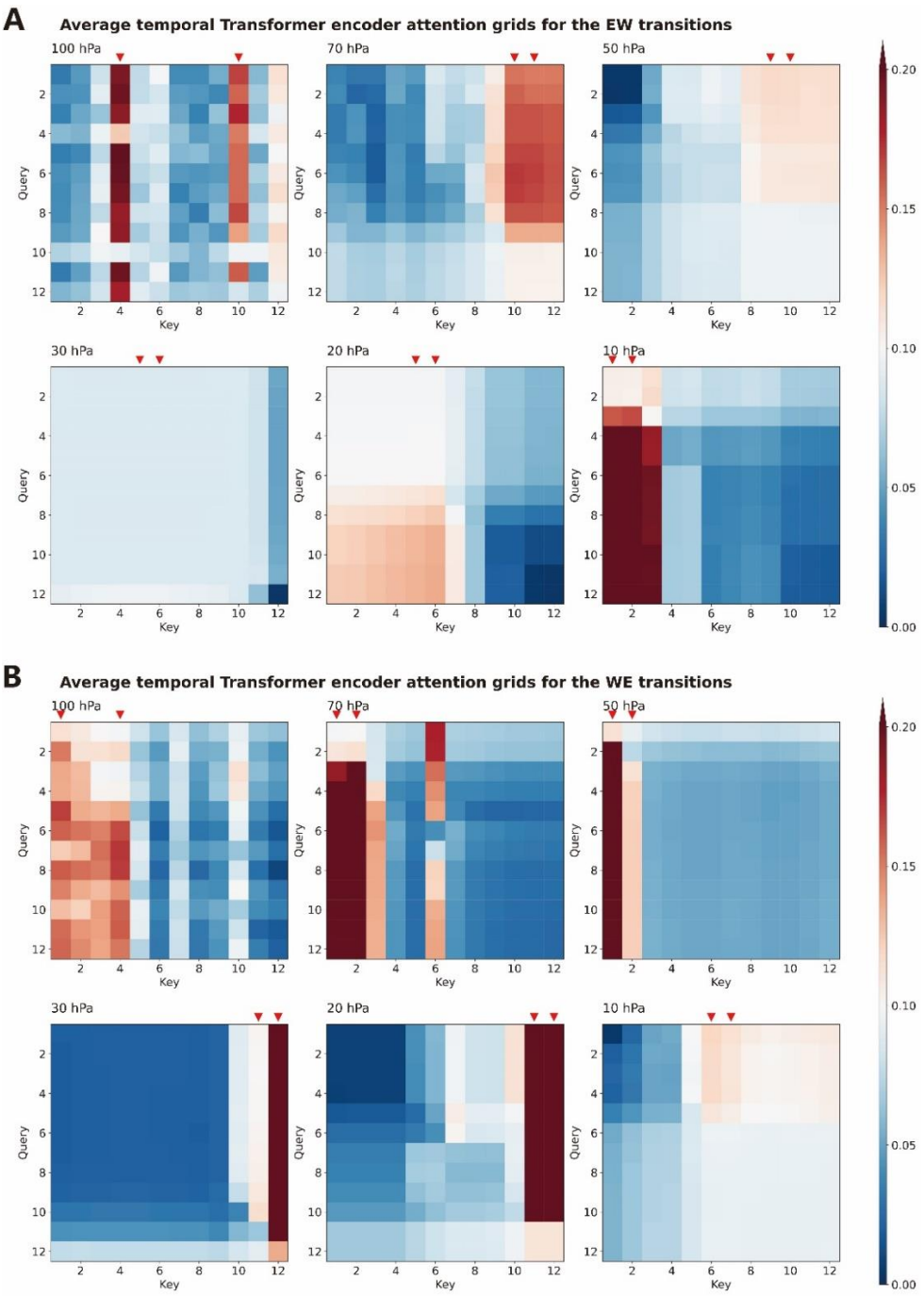
Supplementary Fig. S1 Performance comparison of SETA-DL and dynamical models in simulating the QBO index at 100 hPa over the test dataset.



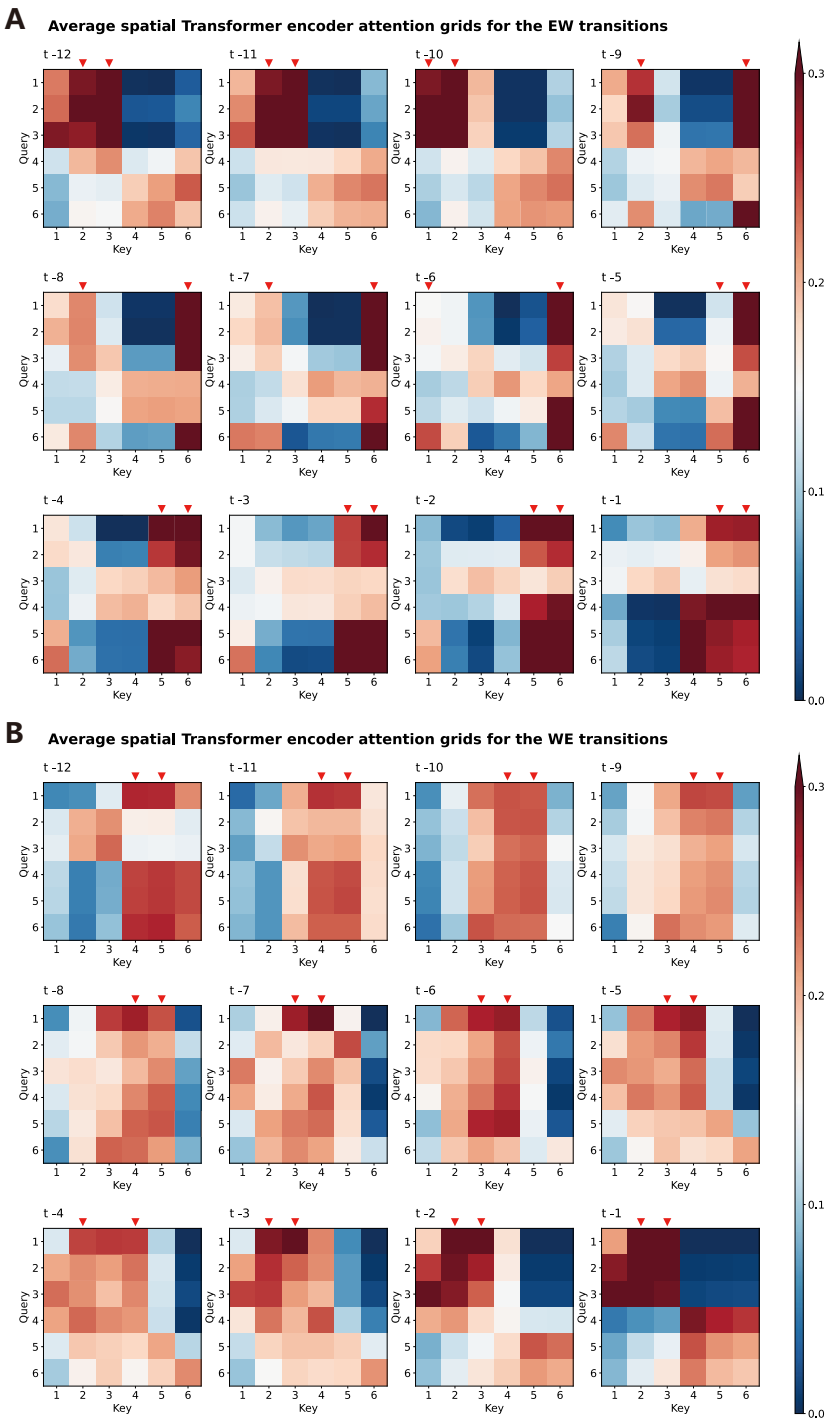
Supplementary Fig. S2 QBO index simulations from SETA-DL (B) and dynamical models (C-G) compared with ERA5 reanalysis (A). Zonal wind is shown by contours, with westerlies (positive values) depicted as solid lines, easterlies (negative values) as dotted lines, and the zero-wind line thickened. The vertical red and blue lines indicate the 2015/16 and 2019/20 QBO disruptions, respectively. Blank regions indicate periods when retrospective simulations are unavailable.



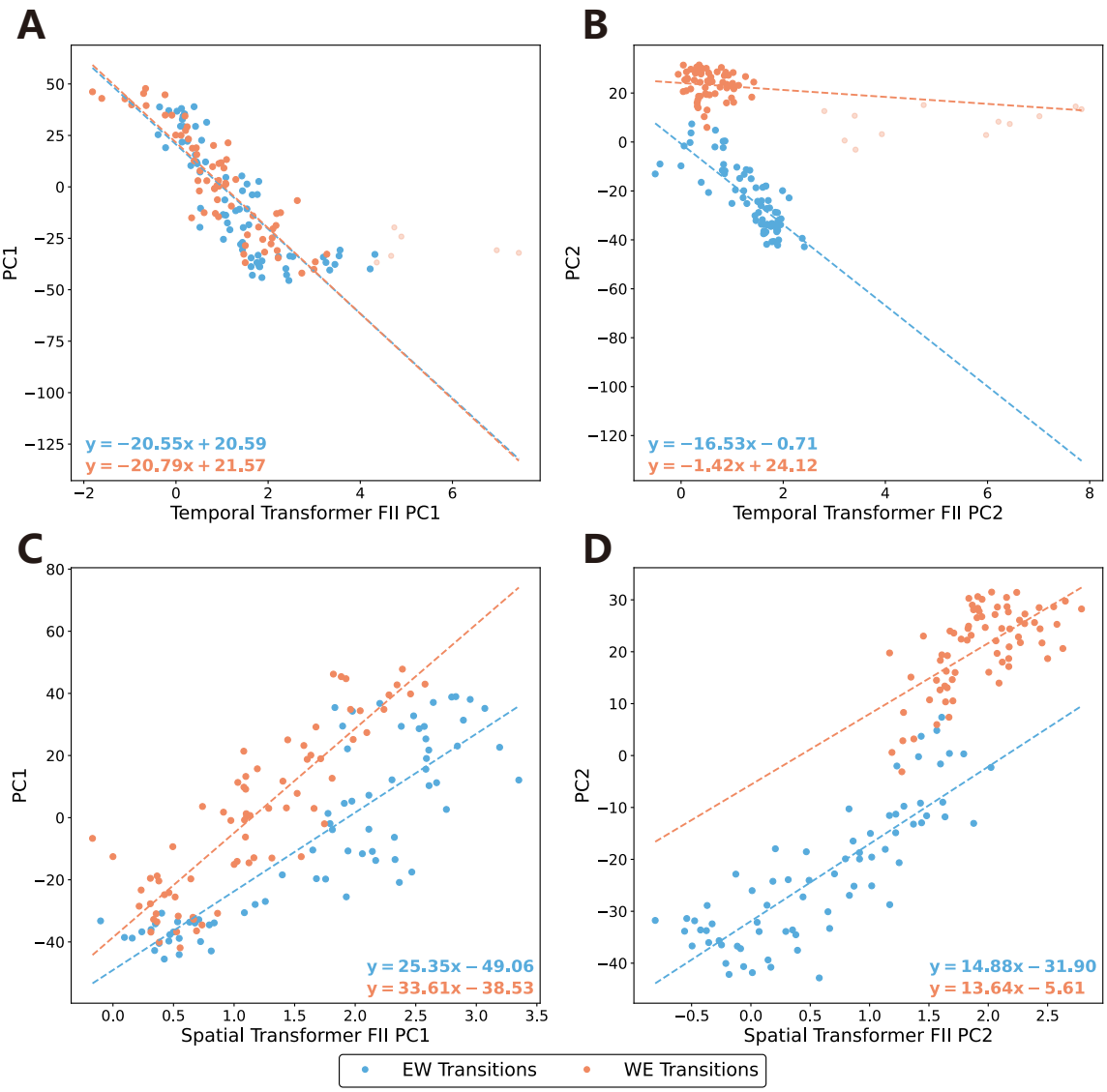
Supplementary Fig. S3 Temporal distribution of easterly-to-westerly (EW) and westerly-to-easterly (WE) QBO transition windows (green shading), with the 30 hPa QBO index shown as a black line. The 12-month periods following each transition window are shown in blue shading. The six EW zonal wind reversals occur in 2001-11, 2004-02, 2006-03, 2008-03, 2010-07, and 2013-02; the six WE reversals occur in 2002-12, 2005-03, 2007-04, 2009-06, 2011-11, and 2014-05.



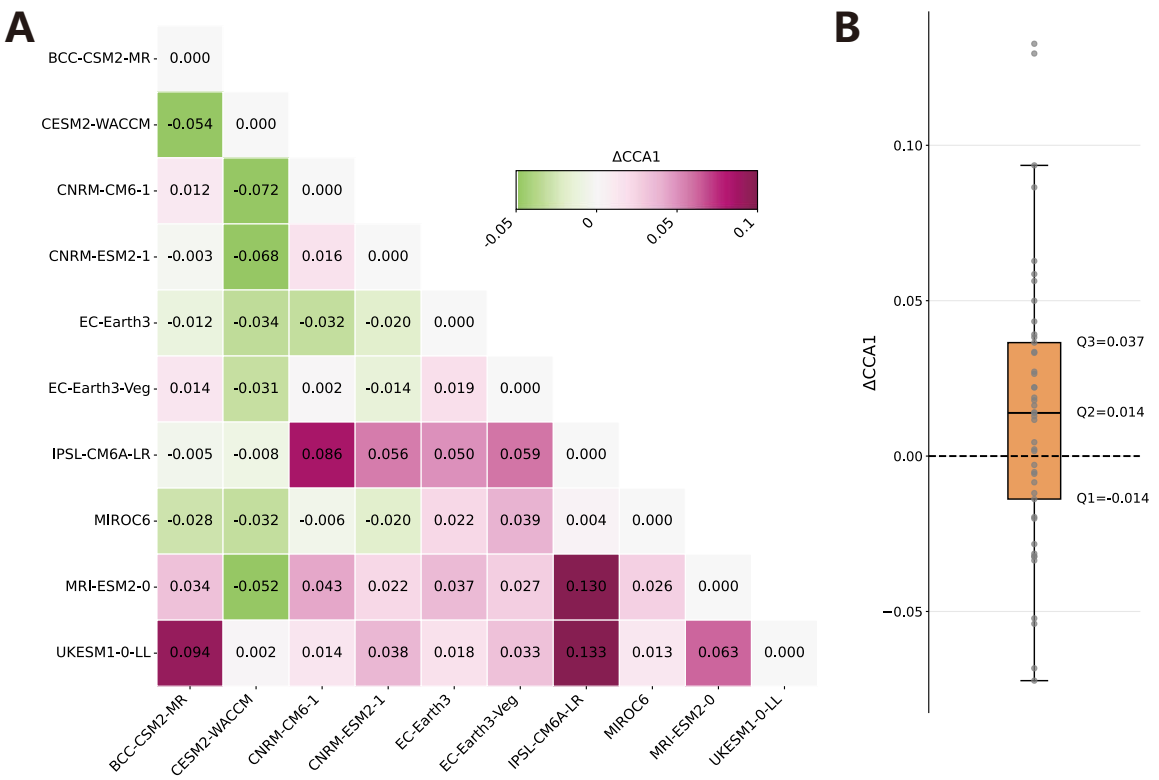
Supplementary Fig. S4 Averages of attention grids from the temporal Transformer encoder. Panels (**A**) and (**B**) show the mean attention patterns computed from easterly-to-westerly (EW) and westerly-to-easterly (WE) QBO transition windows, respectively. Each panel comprises six subfigures, representing the attention grids at different pressure levels. Tick labels on the x- and y-axes indicate positions within the input sequence, with "1" to "12" denoting the first through last month in the input window. Red triangles at the top of each subfigure mark the two columns with the largest summed attention values (i.e., region of interest (ROI)).



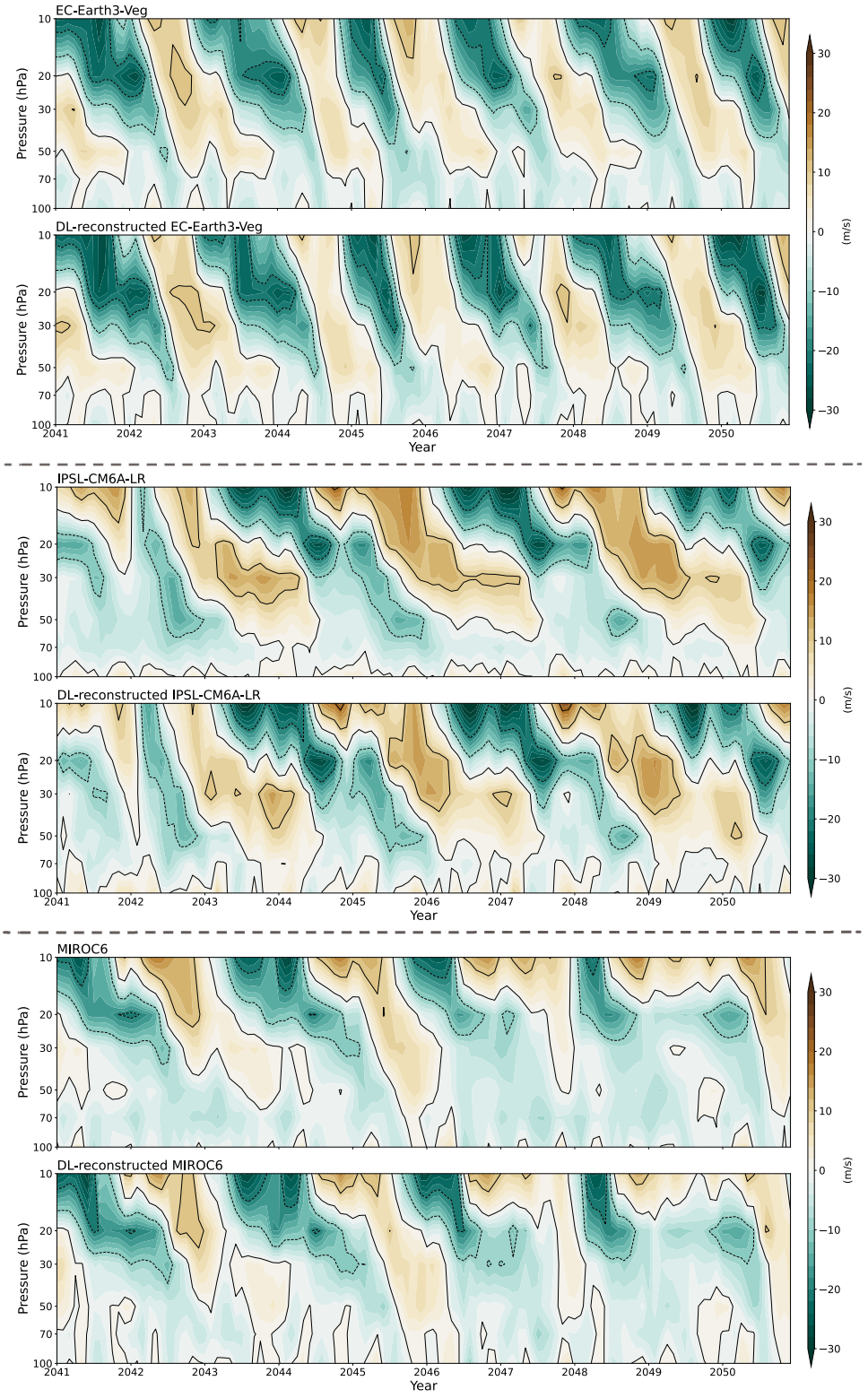
Supplementary Fig. S5 Averages of attention grids from the spatial Transformer encoder. Panels (**A**) and (**B**) show the mean attention patterns computed from EW and WE QBO transition windows, respectively. Each panel comprises twelve subfigures, representing the attention grids at different input months. Tick labels on the x- and y-axes indicate different pressure levels (i.e., "1" to "6" correspond to 100, 70, 50, 30, 20, and 10 hPa pressure levels, respectively). Red triangles at the top of each subfigure mark the two columns with the largest summed attention values.



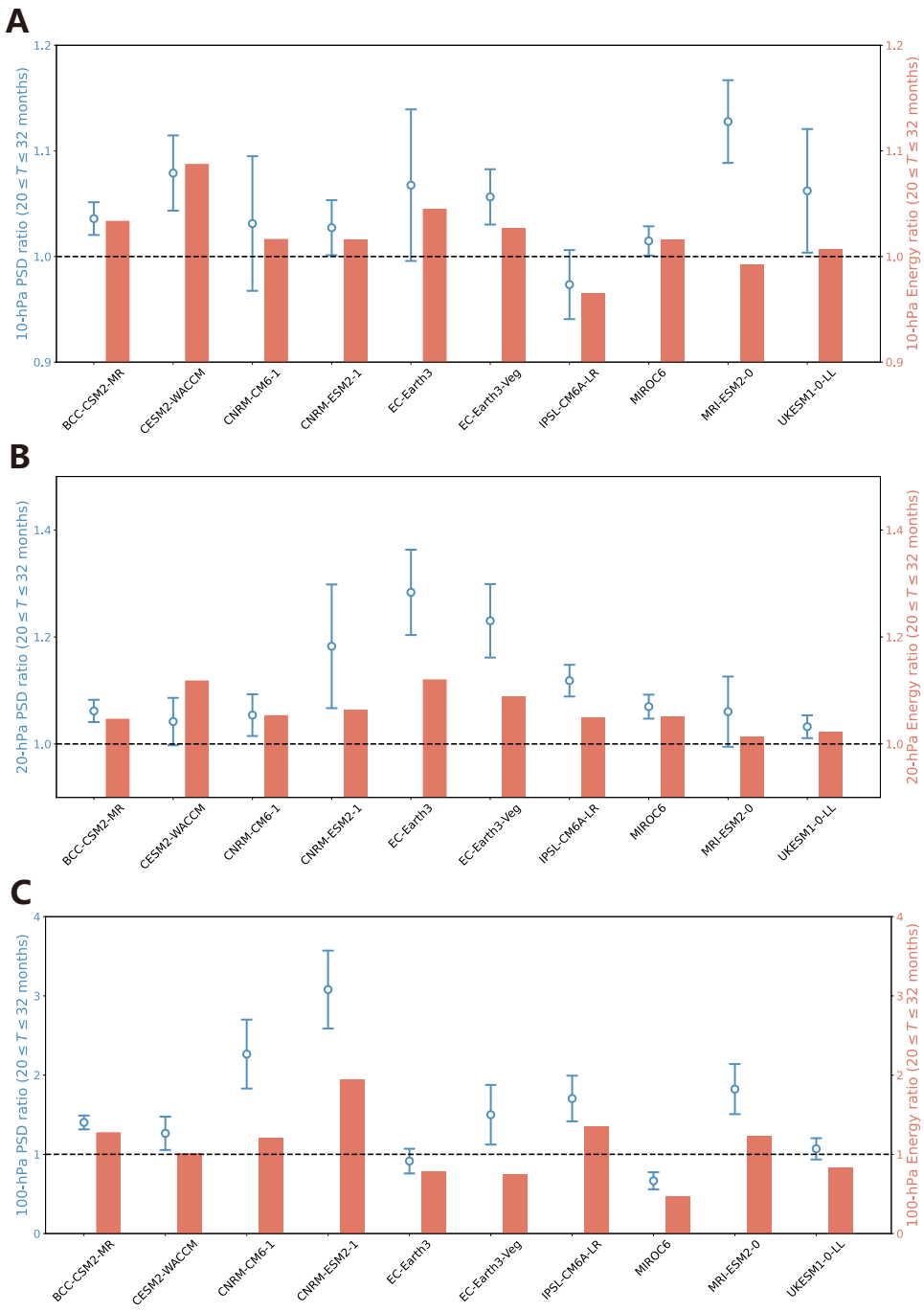
Supplementary Fig. S6 Relationship between the principal components (PCs) of QBO inputs and those of the FIIs. Panels (**A**) and (**C**) show correlations between FII PC1 and QBO PC1, while panels (**B**) and (**D**) show correlations between FII PC2 and QBO PC2. FIIs in (**A-B**) are derived from the temporal encoder, and those in (**C-D**) from the spatial encoder. Linear least-squares fits for EW and WE transitions are shown in blue and orange, respectively. Outlier points, plotted in lighter colors, are excluded from the linear fits (specifically: FII PC1 > 4 for EW transitions in panel (**A**) and FII PC2 > 2 for EW transitions in panel (**B**)).



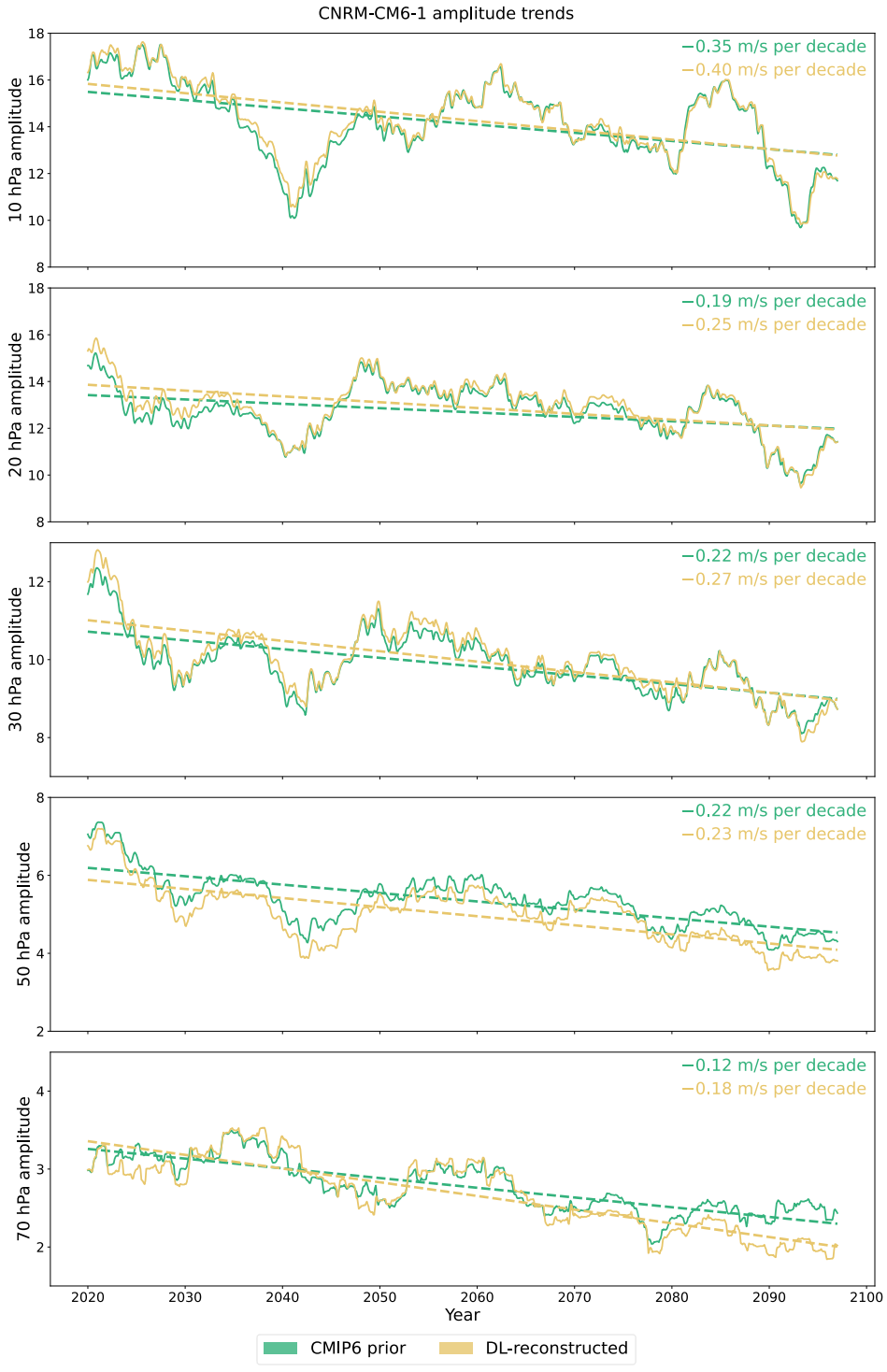
Supplementary Fig. S7 Reduced inter-model uncertainty in climate projections quantified by the first canonical correlation coefficient (CCA1) between all pairs of 10 CMIP6 models. A, Lower triangular matrix of changes in CCA1 ($\Delta\mathrm{CCA1} = \mathrm{CCA1}_{\mathrm{reconstructed}} - \mathrm{CCA1}_{\mathrm{original}}$) between pairs of CMIP6 model projections. Positive values indicate increased similarity (i.e., reduced inter-model spread) after SETA-DL reconstruction. **B,** Boxplot of the 45 unique off-diagonal ΔCCA1 values from panel (**A**), summarizing the overall reduction in inter-model projection diversity.



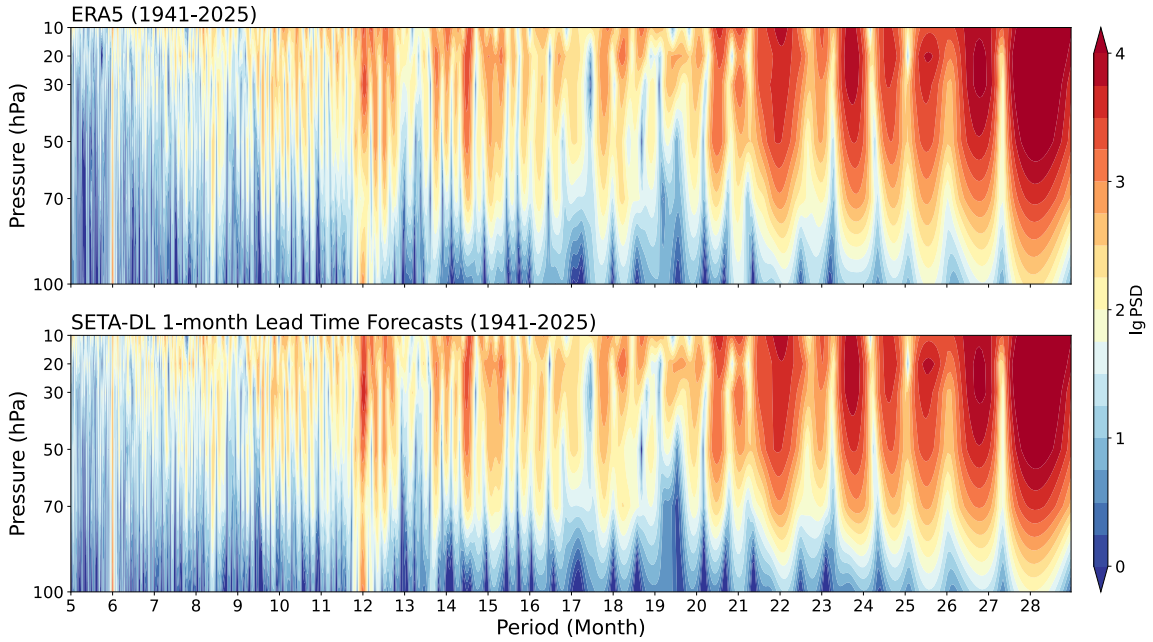
Supplementary Fig. S8 Original and SETA-DL reconstructed SPS585 scenario simulations for 2041-2050. Results from three CMIP6 models (i.e., EC-Earth3-Veg, IPSL-CM6A-LR, and MIROC6) are shown as representative examples.



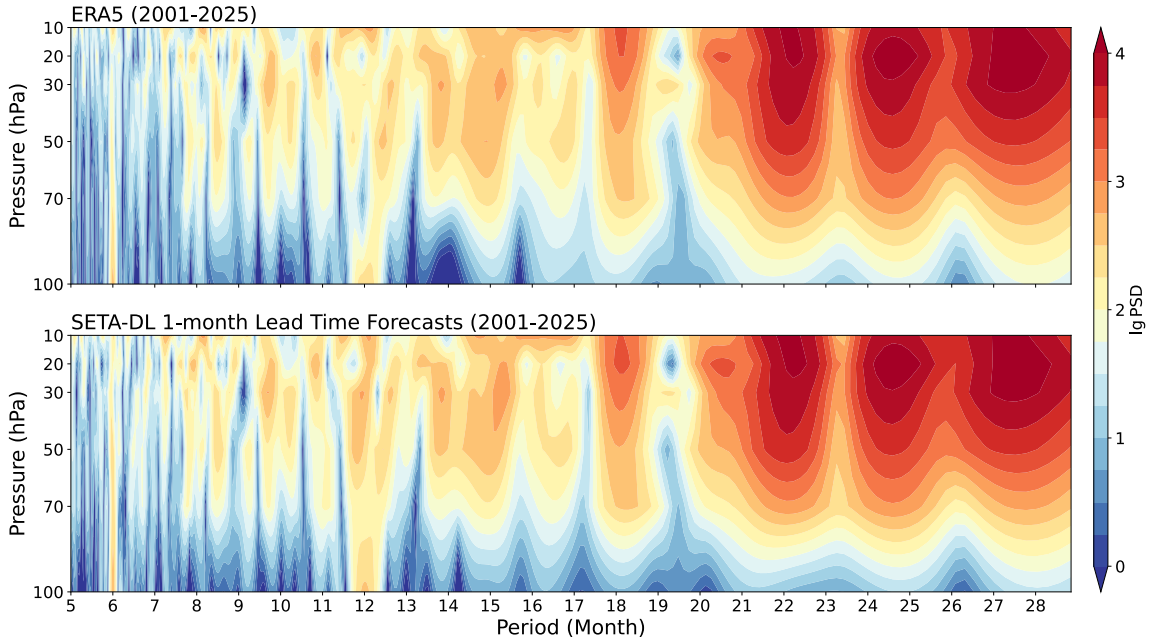
Supplementary Fig. S9 Spectral comparison between the reconstructed and original SSP585 projections at 10, 20, and 100 hPa within the QBO period band (i.e., $20 \leq T \leq 32$ months). The ratios of power spectral density (PSD) are shown with 95% confidence intervals (CIs) around the mean, estimated using a two-sided *t*-test (left y-axis), together with the corresponding total energy ratios integrated over the same band (right y-axis). PSD ratio outliers exceeding 20 are excluded from the analysis and visualization, as these extreme values typically arise from near-zero spectral power in the original simulations at isolated frequencies, leading to artificially inflated ratios.



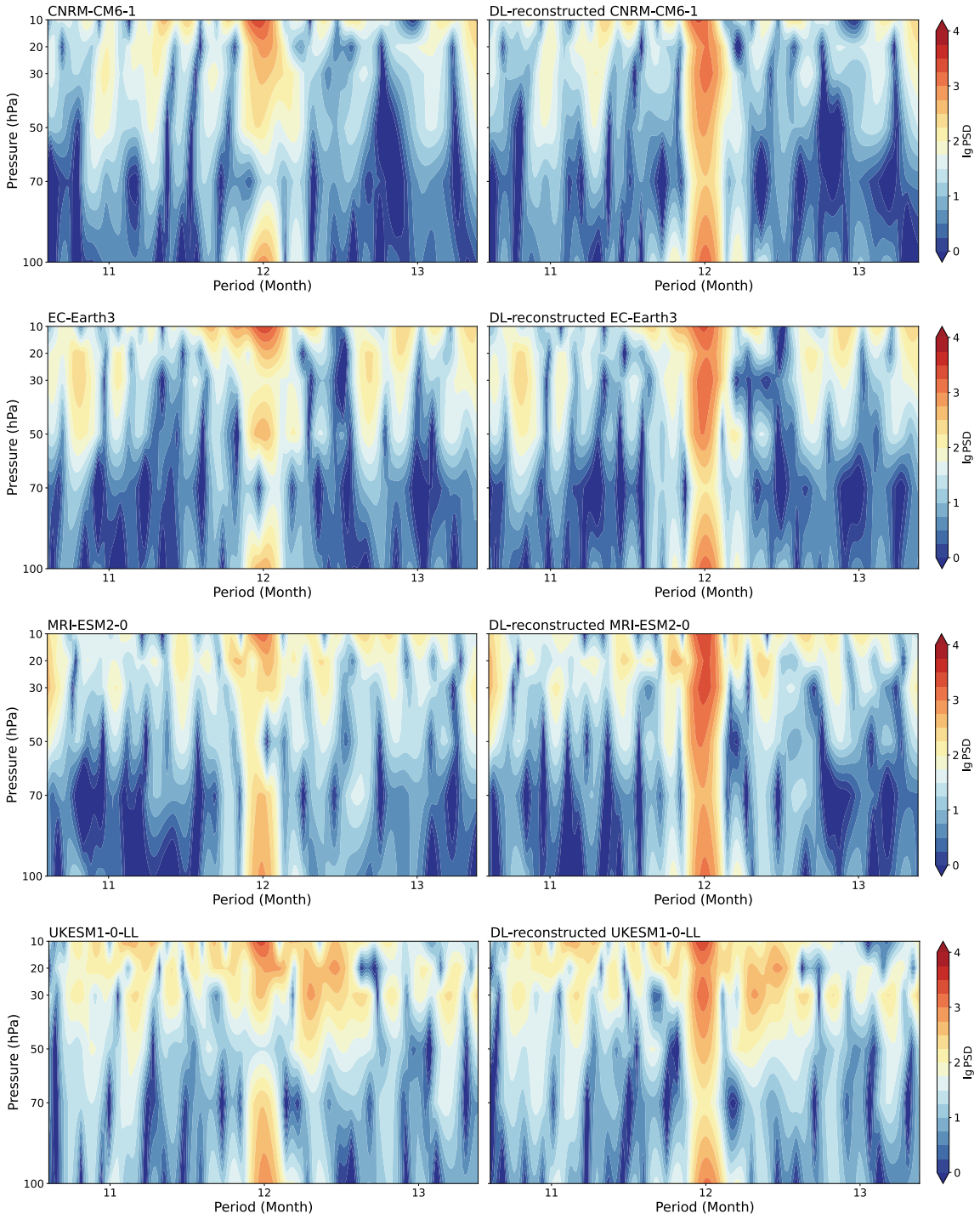
Supplementary Fig. S10. QBO amplitude trends of the original and reconstructed SSP585 projections at levels exhibiting amplitude decay, illustrated for CNRM-CM6-1. Solid lines indicate amplitudes estimated using the Dunkerton-Delisi (DD) method with an 8-year moving window, and points mark the window midpoints. Dashed lines show least-squares linear fits, with the corresponding amplitude trends (i.e., fit slopes) annotated on each plot.



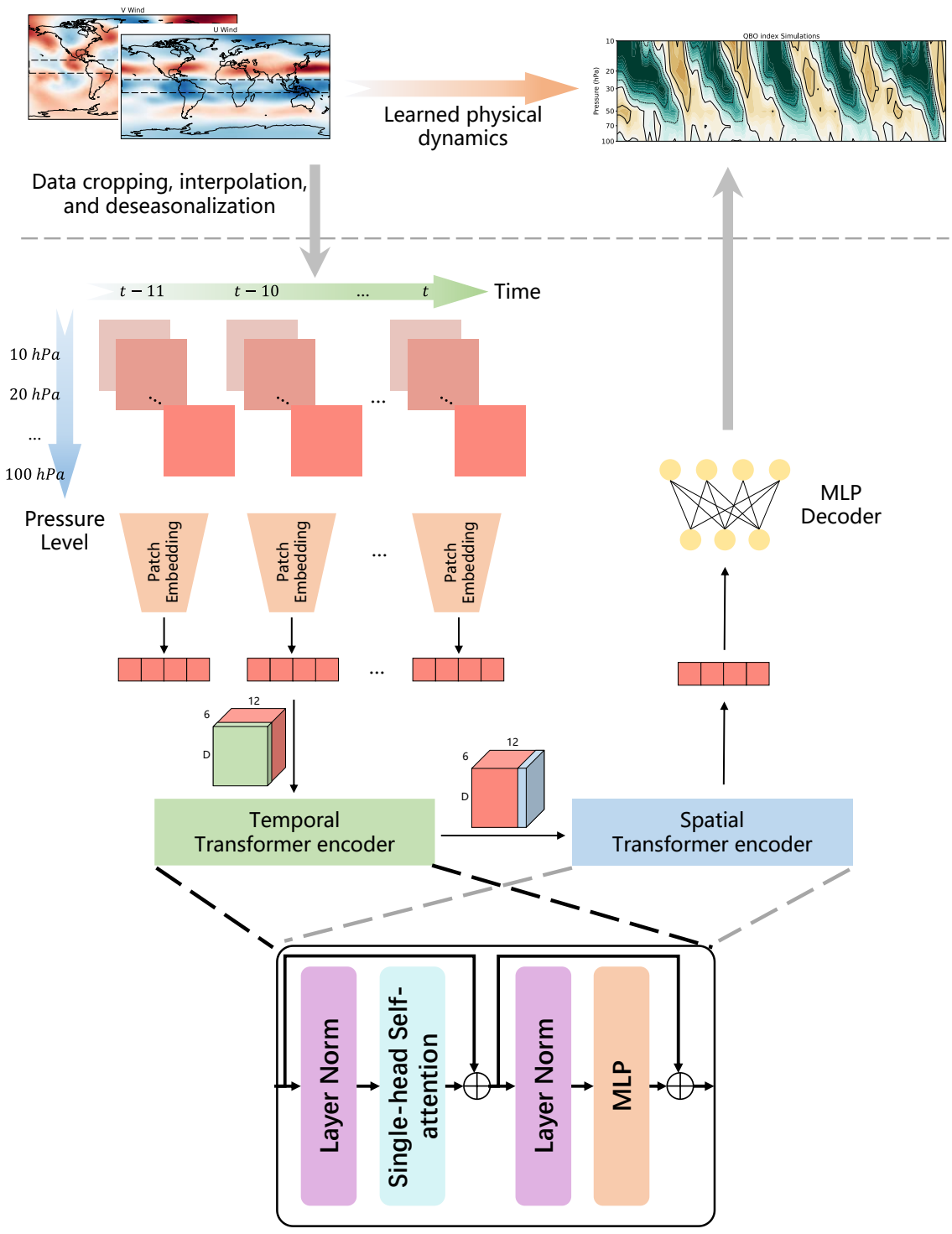
Supplementary Fig. S11 Comparison of PSD between the ERA5-based QBO time series and simulations from SETA-DL, computed over an 85-year segment (1941-2025). Notably, this segment length matches that of the 2016-2100 future projection period. Since the PSD is computed using the fast Fourier transform (FFT), equal record lengths are required to ensure a fair comparison of amplitudes at equivalent frequencies (see "Methods" for PSD computation details).



Supplementary Fig. S12 As in Fig. S11, but for the period 2001 to 2025.



Supplementary Fig. S13 Comparison of the annual oscillation signal in power spectral density (PSD) between the original and SETA-DL reconstructed SSP585 projections. Results from four CMIP6 models (i.e., CNRM-CM6-1, EC-Earth3, MRI-ESM2-0, and UKESM1-0-LL) are displayed as representative examples.



Supplementary Fig. S14 Overall architecture of SETA-DL. SETA-DL comprises four main components: a patch embedding layer, two Transformer encoders, and an MLP decoder. The figure also details the internal structure of the Transformer encoders. The patch embedding layer uses shared weights across all tokens, and *D* denotes the embedding (token) dimension.

Supplementary Table S1. The CMIP6 SPS585 scenario simulations utilized in this study.

CMIP ID	Ensemble Member	Integration Period
BCC-CSM2-MR	r1i1p1f1	2015-2100
CESM2-WACCM	r1i1p1f1	
CNRM-CM6-1	r1i1p1f2	
CNRM-ESM2-1	r1i1p1f2	
EC-Earth3	r4i1p1f1	
EC-Earth3-Veg	r2i1p1f1	
IPSL-CM6A-LR	r1i1p1f1	
MIROC6	r1i1p1f1	
MRI-ESM2-0	r1i1p1f1	
UKESM1-0-LL	r1i1p1f2	

Supplementary Table S2. Architectural specifications and number of trainable parameters for each component of SETA-DL. Activation functions are omitted for clarity. Note: To better learn temporal dynamics, model training is performed with supervision over a 12-month horizon, while inference is restricted to the 1-month ahead simulation by slicing the output sequence.

Model Component	Layer	Data Dimension	Number of Parameters
-	-	$12 \times 6 \times 5 \times 144 \times 2$	-
Patch embedding layer	Conv 2d	$12 \times 6 \times 5 \times 144 \times D$	81
	Linear & average pooling	$12 \times 6 \times D$	
Temporal Transformer encoder	Single-head self-attention (SHSA)	$6 \times (12 \times D)$	244
	Feed forward network (FFN)	$6 \times (12 \times D)$	
Spatial Transformer encoder	SHSA	$12 \times (6 \times D)$	244
	FFN	$12 \times (6 \times D)$	
Decoder	Embedding & Temporal linear	12×6	161
	Slicing	1×6	