Graph network algorithm details

Building a machine learning climate emulator

Harvard Earth and Planetary Science 226, Eli Tziperman

The deterministic model represents each of the 5,810 ocean cells on the 41×170 Pacific grid as a graph node with seven features: standardized SST, D20, and zonal wind stress; two calendar values; and normalized latitude and longitude. Ocean nodes one or two cells apart horizontally or vertically are connected by eight directed relation types that distinguish distance and information arriving from the north, south, east, and west. A linear encoder lifts each node to 64 learned features, and eight residual message-passing layers exchange information among neighboring ocean cells. Finally, a linear decoder maps each node back to three fields and places them onto the grid, producing next-month maps of SST, D20, and zonal wind stress.

1. State and model input

1.1 State, normalization, and calendar

Let $c \in \{1,2,3\}$ denote SST, D20, and zonal wind stress; let $M_{ij} \in \{0,1\}$ be the fixed Pacific mask on the 41×170 grid, $\Omega = \{(i,j) : M_{ij} = 1\}$, and $N = 5810$. From the first $8400 = 700 \times 12$ months, one scalar mean and population standard deviation are stored per field,

$$\mu_c = \frac{1}{8400N} \sum_{t=0}^{8399} \sum_{(i,j)\in\Omega} F_{tcij}, \qquad \sigma_c = \left[\frac{1}{8400N} \sum_{t=0}^{8399} \sum_{(i,j)\in\Omega} (F_{tcij} - \mu_c)^2 \right]^{1/2}. \tag{1}$$

At month t, define $S_{tcij} = (F_{tcij} - \mu_c)/\sigma_c$ on Ω and $S_{tcij} = 0$ outside Ω. Let $m_t \in \{0, \ldots, 11\}$ be the zero-based calendar month. Tensor dimensions below omit the batch axis. During deterministic inference, batch members are evaluated independently using the same learned parameters.

1.2 Ocean-node input

Boolean row-major numbering of Ω creates one node per ocean cell. At node i corresponding to grid cell (a_i, b_i), define

$$\widetilde{\theta}_i = 2\frac{\theta_{a_i} - \theta_{\min}}{\theta_{\max} - \theta_{\min}} - 1, \quad \widetilde{\lambda}_i = 2\frac{\lambda_{b_i} - \lambda_{\min}}{\lambda_{\max} - \lambda_{\min}} - 1. \tag{2}$$

Here θ and λ are latitude and longitude; their normalized values are fixed positional encodings. The seven node features are

$$x_i = \left[S_{t1,a_ib_i}, S_{t2,a_ib_i}, S_{t3,a_ib_i}, \sin(2\pi m_t/12), \cos(2\pi m_t/12), \widetilde{\theta}_i, \widetilde{\lambda}_i \right] \in \mathbb{R}^7. \tag{3}$$

The graph never includes land nodes, so it needs neither a land-mask feature nor land messages.

2. Network building blocks

2.1 Graph connections between ocean cells

The Pacific grid is represented by a *graph*: its "nodes" are the 5,810 ocean cells, and its "edges" connect ocean cells one or two grid steps to their east/west or north/south. The graph is *directed* because every edge has a source and target. The edges are *typed*: each is assigned a direction and distance label, with a different learned transformation for each label.

Each neighboring pair A, B contributes two directed edges, $A \to B$ and $B \to A$. There is no diagonal edge or edge across land; a two-step edge requires a wet midpoint. The mask yields 22,422 pairs and 44,844 directed edges. Relation IDs below are listed for distances $d = 1, 2$, respectively:

r	source $\to$ target	incoming information at target
0, 4	latitude index $a \to a + d$	from the lower-index (southern) cell d steps away
1, 5	latitude index $a + d \to a$	from the higher-index (northern) cell d steps away
2, 6	longitude index $b \to b + d$	from the lower-index (western) cell d steps away
3, 7	longitude index $b + d \to b$	from the higher-index (eastern) cell d steps away

For each relation r, let $j_r(i)$ be the unique node that sends information to node i from that direction and distance, when such a neighbor exists. Let $I_{ir} = 1$ when it exists and $I_{ir} = 0$ otherwise.

2.2 Fully connected encoder and activation

A fully connected layer shared across all ocean nodes maps the seven input values at each node to 64 learned features, followed by GELU activation:

$$h_i^{(0)} = \mathrm{GELU}\left(x_i (W^{\mathrm{enc}})^\top + b^{\mathrm{enc}} \right), \qquad W^{\mathrm{enc}} \in \mathbb{R}^{64\times 7},\ b^{\mathrm{enc}} \in \mathbb{R}^{64}. \tag{4}$$

Here $h_i^{(0)}$ is the lifted representation of node i; the superscript (0) indicates that it is the input to the first message-passing layer. The exact GELU is applied componentwise:

$$\mathrm{GELU}(u) = u\Phi(u) = \frac{u}{2}\left[1 + \mathrm{erf}\left(\frac{u}{\sqrt{2}} \right) \right]. \tag{5}$$

2.3 One message-passing layer and its residual update

For message-passing layer $\ell \in \{1, \dots, 8\}$, use a separate full 64×64 matrix $W_r^{(\ell)}$ for each of the eight direction–distance relations $r = 0, \dots, 7$, plus a 64×64 root weight matrix $W_{\text{self}}^{(\ell)}$ and bias $b^{(\ell)}$. The code combines these nine matrix blocks in one affine map from $9 \times 64 = 576$ to 64 features. The root weight matrix lets node i's own ("self") features contribute to the update, while the bias supplies a learned offset. Each incoming neighbor is transformed by the matrix for its direction and distance. Adding the transformed node, each available transformed neighbor, and the bias gives $g_i^{(\ell)} \in \mathbb{R}^{64}$, the intermediate 64-feature output of the message-passing calculation for node (grid cell) i, before GELU activation and the residual addition:

$$g_i^{(\ell)} = h_i^{(\ell-1)} (W_{\text{self}}^{(\ell)})^\top + b^{(\ell)} + \sum_{r=0}^{7} I_{ir}\, h_{j_r(i)}^{(\ell-1)} (W_r^{(\ell)})^\top, \tag{6}$$

where the term is defined as zero when $I_{ir} = 0$. Thus the sum contains at most eight transformed neighbor contributions. The layer applies GELU to $g_i^{(\ell)}$ and adds node i's previous features $h_i^{(\ell-1)}$. This residual addition gives the layer output:

$$h_i^{(\ell)} = h_i^{(\ell-1)} + \text{GELU}\left(g_i^{(\ell)}\right), \qquad \ell = 1, \dots, 8. \tag{7}$$

Therefore one layer communicates up to two cardinal grid steps, and after eight layers $h_i^{(8)}$ can depend on wet nodes reachable within eight graph edges (up to sixteen grid steps). The learned self term is inside $g_i^{(\ell)}$ and the identity residual is outside it; both are present. There is no batch/layer normalization and no dropout.

For a batch of states, the implementation keeps a separate batch axis and uses the same neighbor-index table, with missing neighbors gathered from an appended zero node.

3. Layer-by-layer evaluation

3.1 Network evaluation and output

Apply the encoder and eight message-passing layers above, then decode each processed node directly to the three next-month standardized fields,

$$q_i = h_i^{(8)} (W^{\text{dec}})^\top + b^{\text{dec}}, \qquad W^{\text{dec}} \in \mathbb{R}^{3 \times 64}, \quad b^{\text{dec}} \in \mathbb{R}^3. \tag{8}$$

The program first allocates $Y_t \in \mathbb{R}^{3 \times 41 \times 170}$ as all zeros, then places each node's three decoded values back into the ocean grid cell represented by that node:

$$Y_{t,c,a_i,b_i} = q_{ic} \quad (i = 1{:}5810), \qquad Y_{t,cab} = 0 \quad \text{when } M_{ab} = 0. \tag{9}$$

The number of ocean (unmasked) grid cells is 5,810. The three decoded values for each node (grid cell)—SST, D20, and τ_x—are placed back in that node's original grid cell.

3.2 Dimensions and parameter accounting

stage	input	exact map	output	parameters
encode	5810×7	affine $7 \to 64$, GELU	5810×64	512
each of 8 processors	5810×64	8 relation matrices + self + bias	5810×64	36,928
decode	5810×64	affine $64 \to 3$	5810×3	195
place on grid	5810×3	ocean rows into zero grid	$3 \times 41 \times 170$	0
			total trainable parameters	**296,131**

4. Deterministic monthly recurrence

Starting from a standardized state S_t, construct the model input described in Section 1 and evaluate the network to obtain $Y_t = f_\theta(S_t, m_t; M)$, with no output activation. During a deterministic free run, impose the ocean mask and advance the calendar:

$$S_{t+1,cij} = M_{ij} Y_{t,cij}, \qquad m_{t+1} = (m_t + 1) \bmod 12. \tag{10}$$

Rebuild the model input from the new state and calendar month and repeat the complete network. Saved ocean output is converted to physical units field by field:

$$\widehat{F}_{t+1,cij} = \sigma_c S_{t+1,cij} + \mu_c \quad ((i,j) \in \Omega). \tag{11}$$

Land is written as missing (NaN). The only state carried between forecast months is the three-channel standardized prediction; no hidden features are carried forward. The calendar advances analytically, while the mask and grid remain fixed.

5. Architecture at a glance

