GraphCast

Presented by Martynas Vaznonis

Traditional weather prediction

Numerical weather prediction

- HRES
- Ensemble method

Atmospheric Equations $[\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{V} \bullet \nabla]$

Motion $$\frac{D\vec{V}}{Dt} - f\hat{k} \times \vec{V} = -\frac{1}{\rho}\nabla P - g\hat{k} + \nu\nabla^2\vec{V} \qquad \vec{V} = \hat{i}u + \hat{j}v + \hat{k}w$$

Mass continuity $$\frac{\partial \rho}{\partial t} + \nabla \bullet (\rho\vec{V}) = 0$$

Thermodynamic $$c_p \frac{DT}{Dt} - \frac{1}{\rho}\frac{Dp}{Dt} = L(c-e) - \nabla \bullet \vec{F}_{rad} + k\nabla^2 c_p T + \mathfrak{D}$$

Or: $$\frac{DS}{Dt} - gw - \frac{1}{\rho}\frac{Dp}{Dt} = L(c-e) + k\nabla^2 T + \mathfrak{D} \qquad S = c_p T + \Phi \qquad \frac{D\Phi}{Dt} = gw$$

Vapour $$\frac{Dq_v}{Dt} = e - c$$

Condensed water $$\frac{Dq_w}{Dt} = c - e - \mathrm{Pr}$$

Equation of State (ideal gas) $$P = \rho R T_v \qquad T_v \cong T(1 + .61q_v - q_w)$$

Motivation

Benefits of machine learning

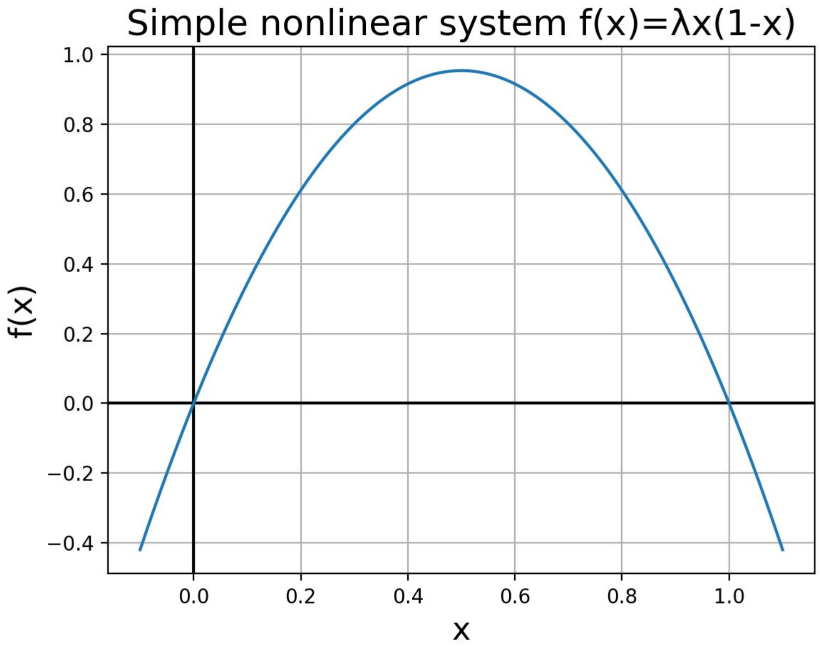
- Can capture patterns not easily represented by equations
- Can make use of historical data

Weaknesses of numerical methods

- Relatively weak for some tasks
  - Sub-seasonal heat wave prediction
  - Precipitation nowcasting from radar images
- Slow

Unpredictability

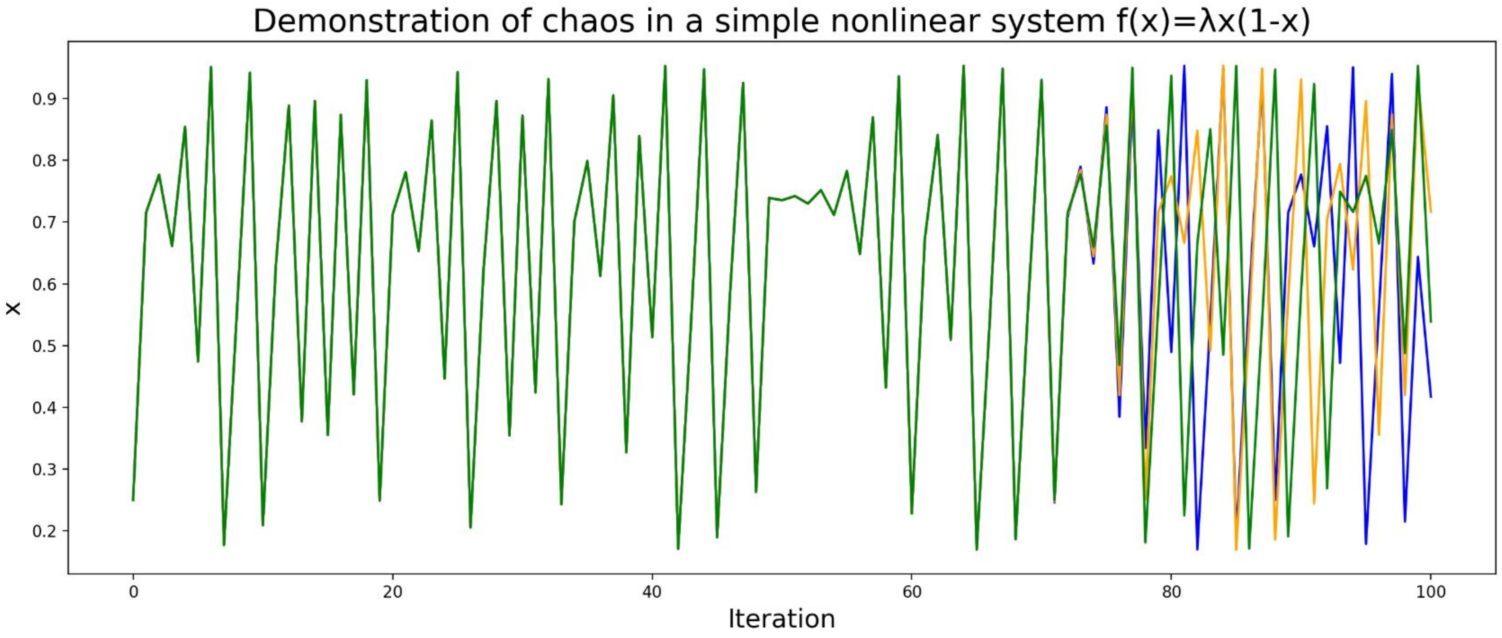
- Weather is nonlinear
- Nonlinear systems are chaotic



- $x'=f(x)$
- $x''=f(x')$
- $x_0=0.25$

Unpredictability

Absolute difference of ~$5*10^{-17}$



Unpredictability

- Extremely small errors amplify and double every 5 days
- Max 2 week lead time forecasting
- Machine learning might help

Atmospheric Equations $\left[\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{V} \bullet \nabla\right]$

Motion
$$\frac{D\vec{V}}{Dt} - f\hat{k} \times \vec{V} = -\frac{1}{\rho}\nabla P - g\hat{k} + \nu\nabla^2\vec{V} \qquad \vec{V} = \hat{i}u + \hat{j}v + \hat{k}w$$

Mass continuity
$$\frac{\partial \rho}{\partial t} + \nabla \bullet \left(\rho\vec{V}\right) = 0$$

Thermodynamic
$$c_p \frac{DT}{Dt} - \frac{1}{\rho}\frac{Dp}{Dt} = L(c-e) - \nabla \bullet \vec{F}_{rad} + k\nabla^2 c_p T + \mathfrak{D}$$

Or:
$$\frac{DS}{Dt} - gw - \frac{1}{\rho}\frac{Dp}{Dt} = L(c-e) + k\nabla^2 T + \mathfrak{D} \qquad S = c_p T + \Phi \qquad \frac{D\Phi}{Dt} = gw$$

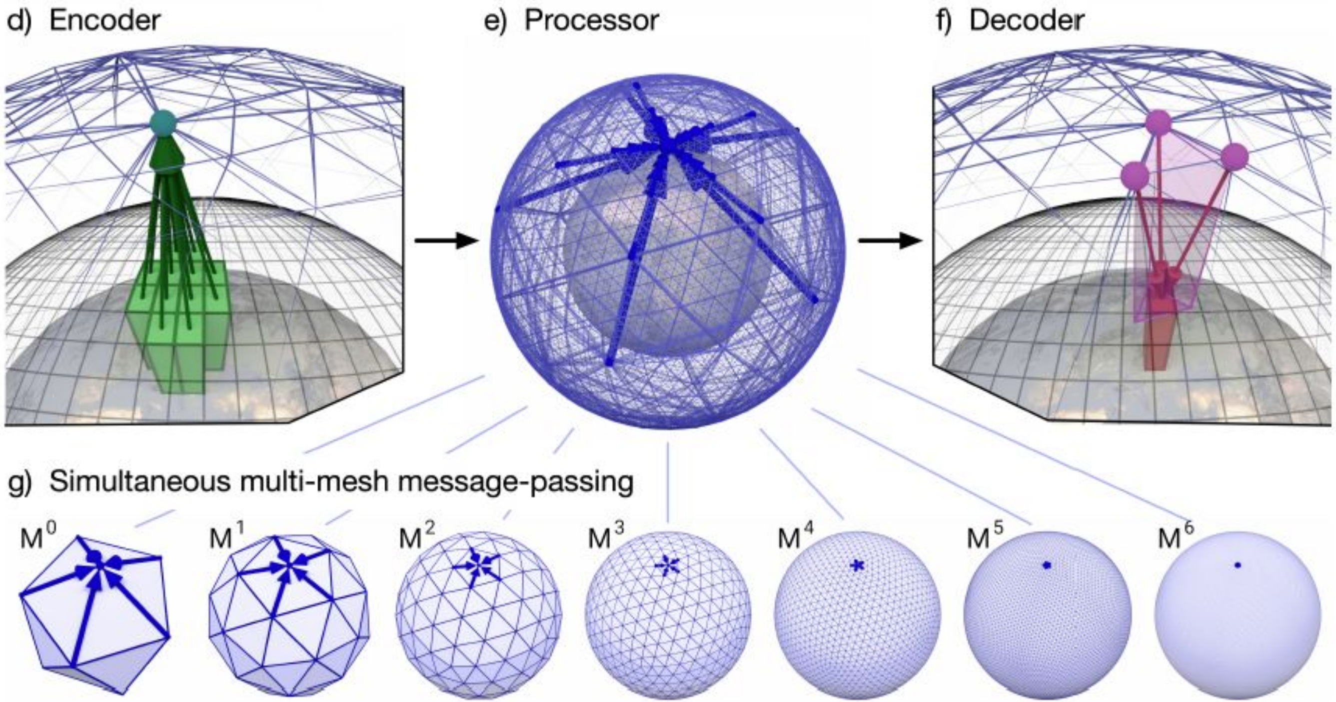
Vapour
$$\frac{Dq_v}{Dt} = e - c$$

Condensed water
$$\frac{Dq_w}{Dt} = c - e - \Pr$$

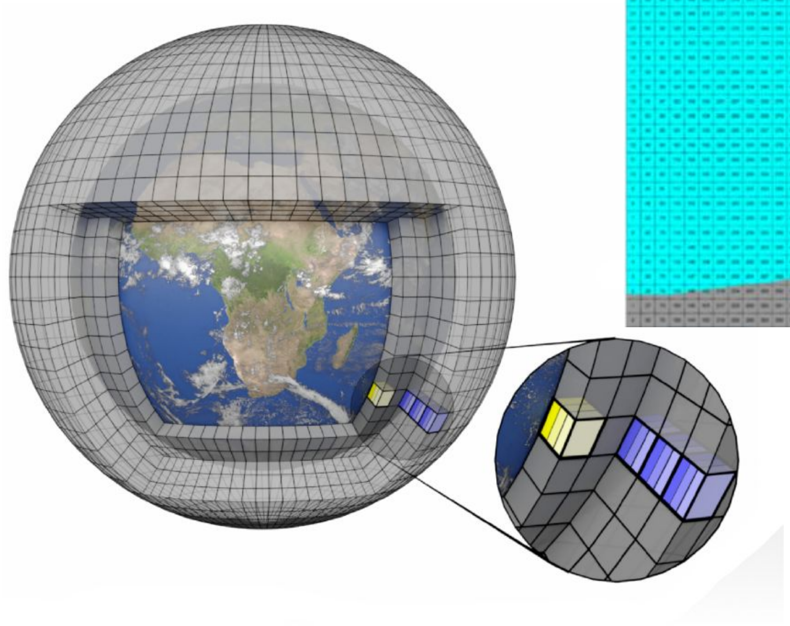
Equation of State (ideal gas)
$$P = \rho R T_v \qquad T_v \cong T(1 + .61 q_v - q_w)$$

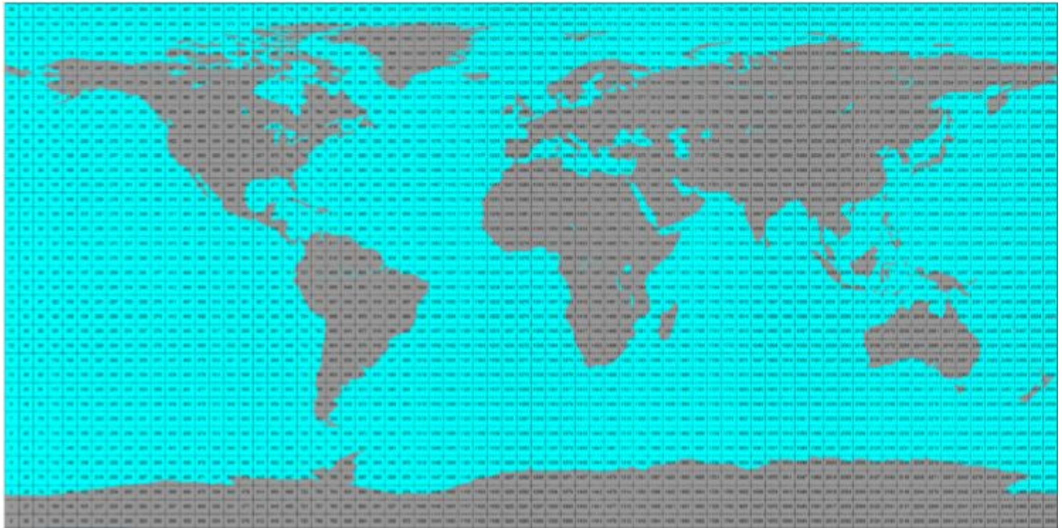
GraphCast overview

- Grid
- Mesh
- Encoder
- Processor
- Decoder



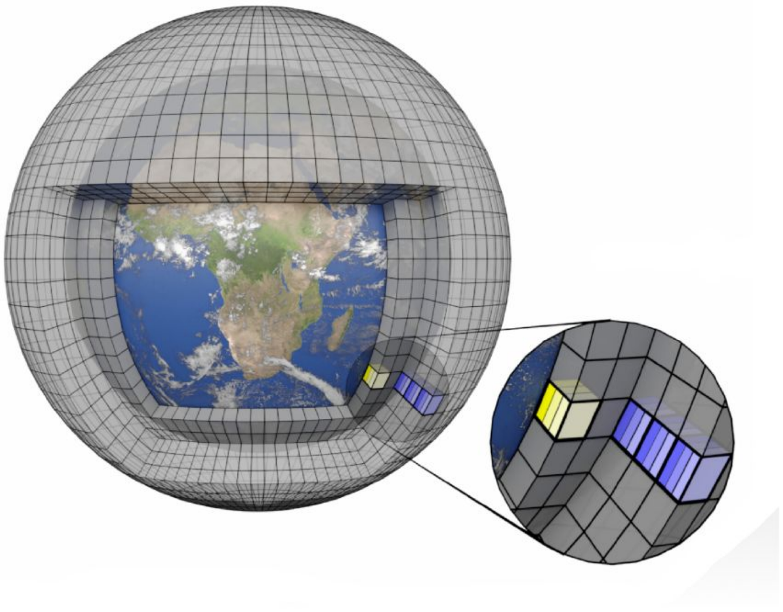
Grid



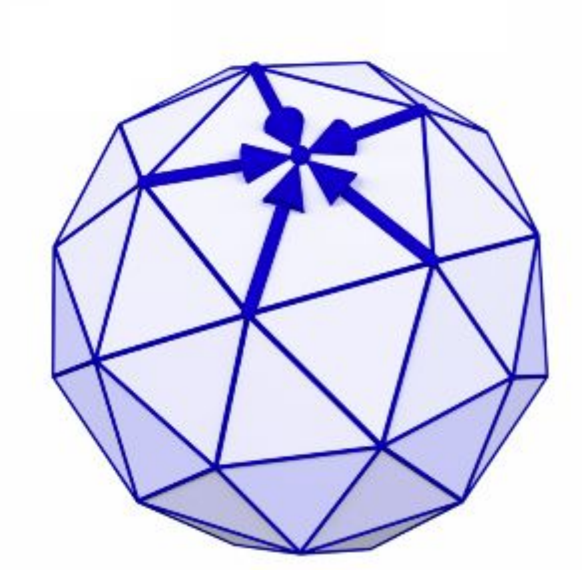


Grid

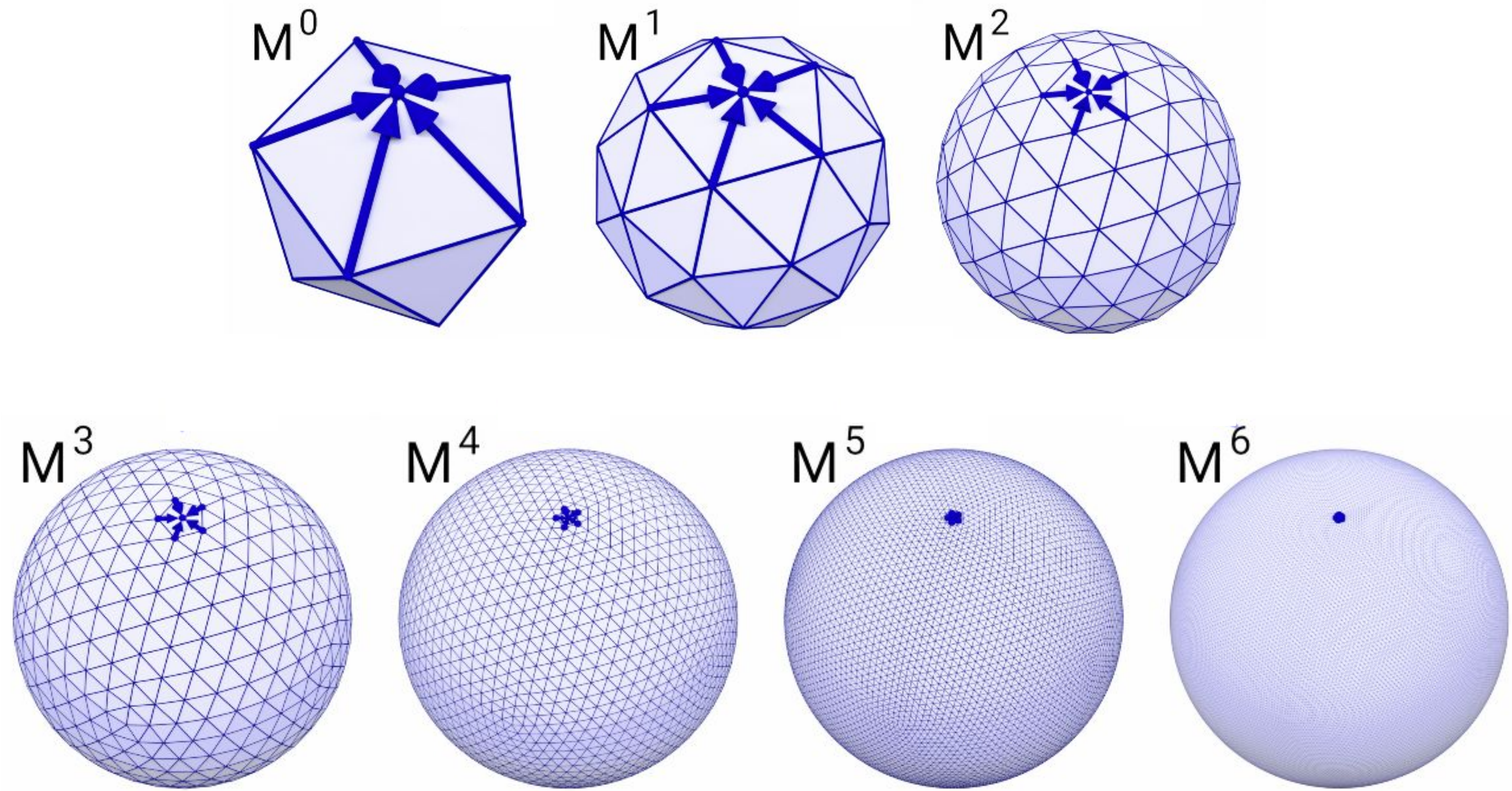
- Ground variables
  - 2-meter temperature
  - Total precipitation
  - 5 total
- Atmospheric variables
  - Wind components
  - 6 per pressure level
  - 37 total pressure levels
- Other features
  - Forcing terms (5)
  - Constants (5)



Mesh

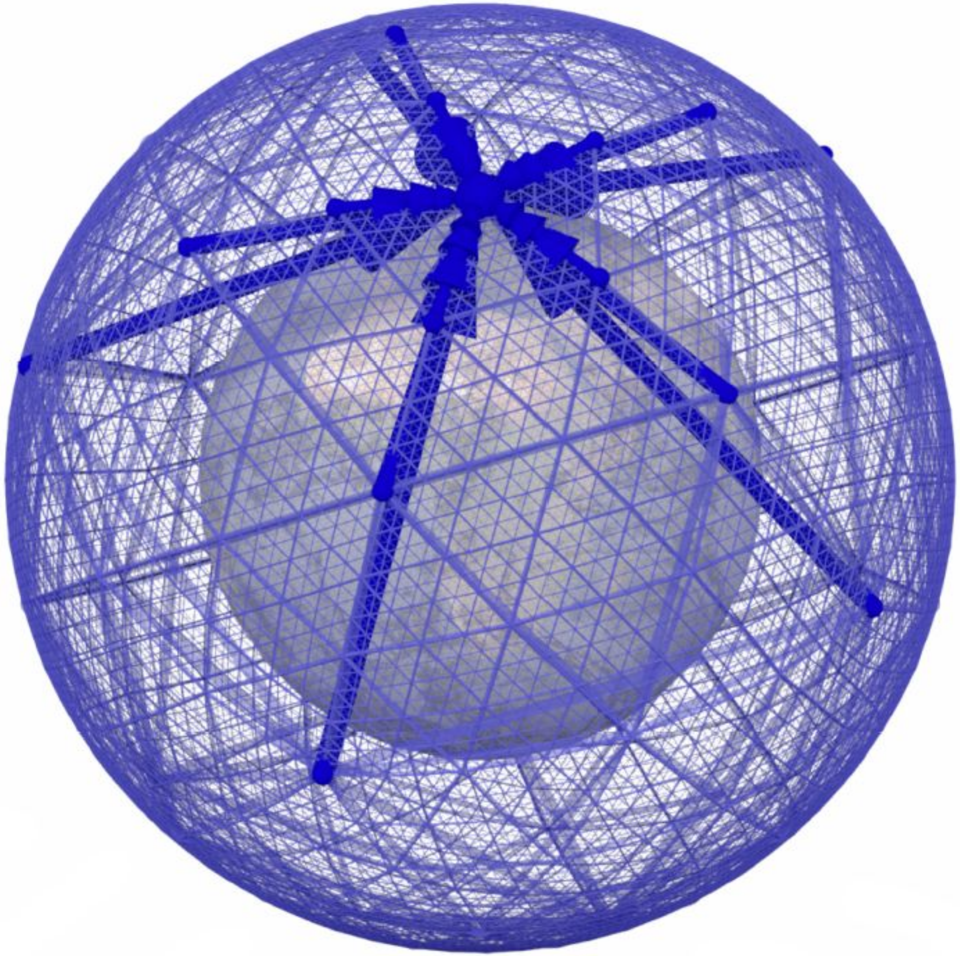


Mesh



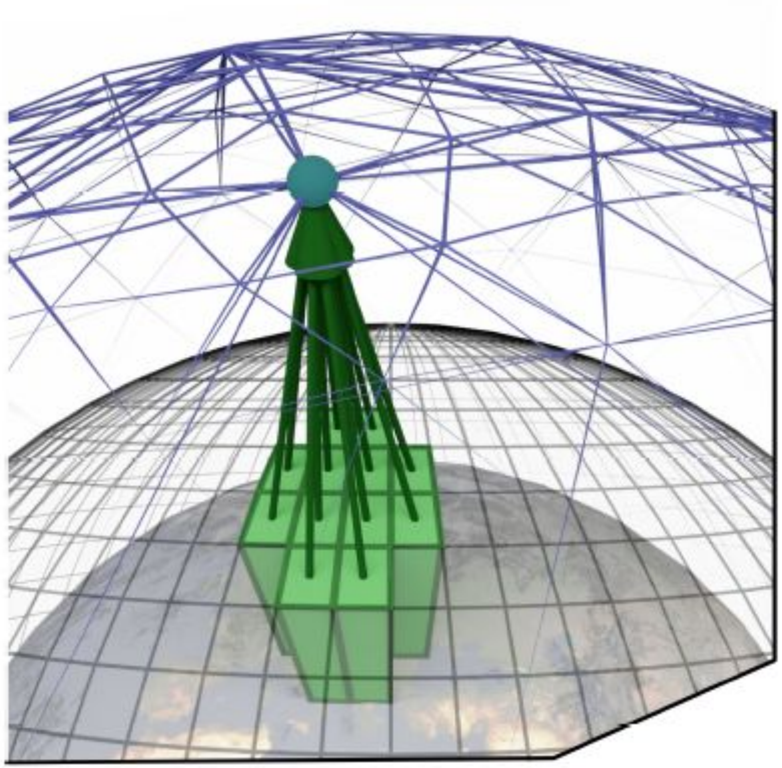
Mesh

- Defines the processor graph
- Homogeneous over the Earth



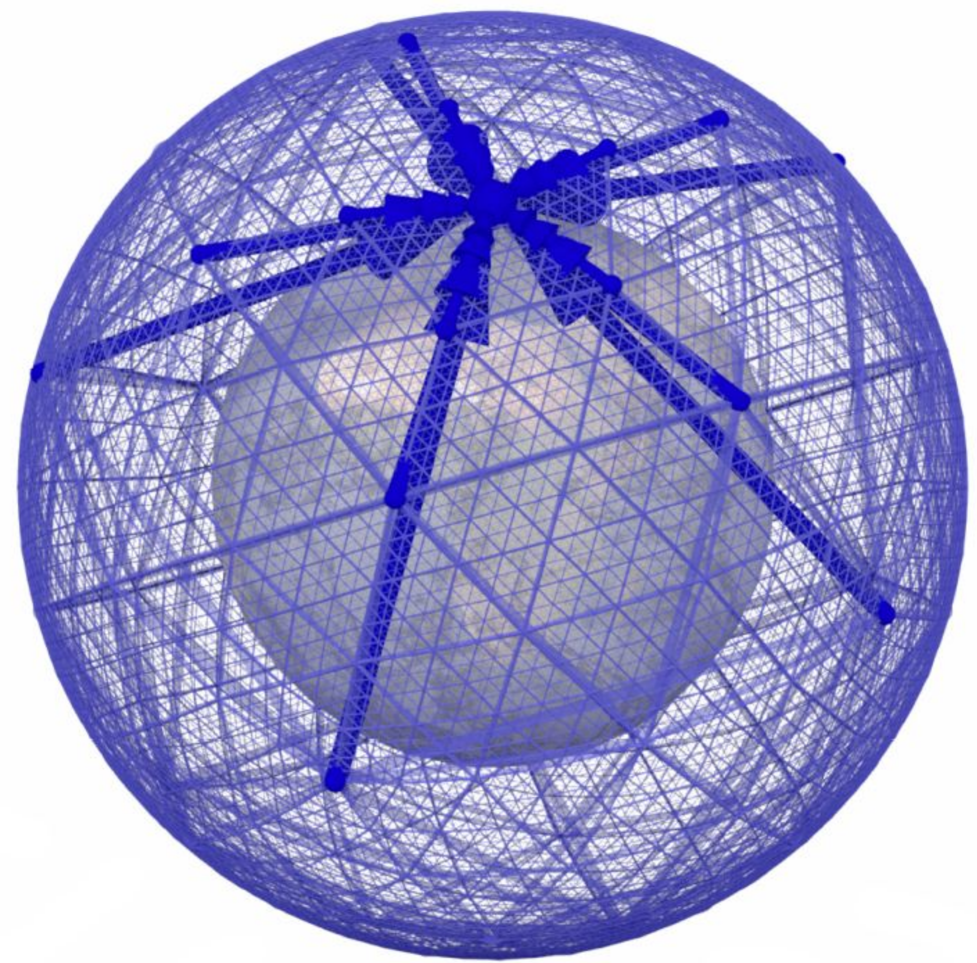
Encoder

- Grid to mesh
- Defines a bipartite graph
- Radius hyperparameter
- Edges have 4 features



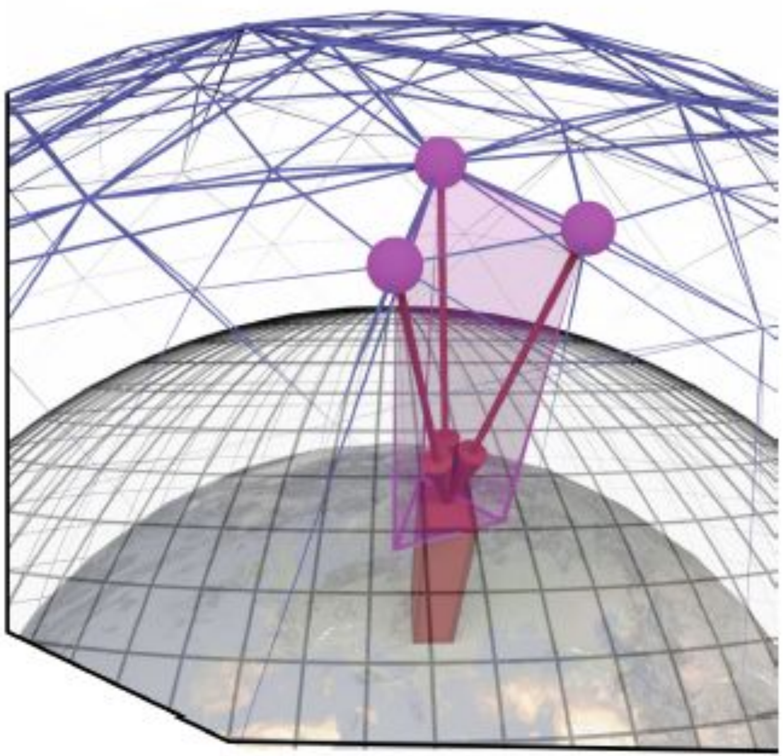
Processor

- Message passing
- 16 GNN layers
- Edge features



Decoder

- Similar to the encoder
- Edge features
- Predicts residual update to current state



Prediction pipeline

$$\mathcal{G}(\mathcal{V}^{\mathrm{G}}, \mathcal{V}^{\mathrm{M}}, \mathcal{E}^{\mathrm{M}}, \mathcal{E}^{\mathrm{G2M}}, \mathcal{E}^{\mathrm{M2G}})$$

Prediction pipeline: embedding

$$\mathbf{v}_i^{\mathrm{G}} = \mathrm{MLP}_{\mathcal{V}^{\mathrm{G}}}^{\mathrm{embedder}}(\mathbf{v}_i^{\mathrm{G,features}})$$

$$\mathbf{v}_i^{\mathrm{M}} = \mathrm{MLP}_{\mathcal{V}^{\mathrm{M}}}^{\mathrm{embedder}}(\mathbf{v}_i^{\mathrm{M,features}})$$

$$\mathbf{e}_{v_{\mathrm{s}}^{\mathrm{M}} \to v_{\mathrm{r}}^{\mathrm{M}}}^{\mathrm{M}} = \mathrm{MLP}_{\mathcal{E}^{\mathrm{M}}}^{\mathrm{embedder}}(\mathbf{e}_{v_{\mathrm{s}}^{\mathrm{M}} \to v_{\mathrm{r}}^{\mathrm{M}}}^{\mathrm{M,features}})$$

$$\mathbf{e}_{v_{\mathrm{s}}^{\mathrm{G}} \to v_{\mathrm{r}}^{\mathrm{M}}}^{\mathrm{G2M}} = \mathrm{MLP}_{\mathcal{E}^{\mathrm{G2M}}}^{\mathrm{embedder}}(\mathbf{e}_{v_{\mathrm{s}}^{\mathrm{G}} \to v_{\mathrm{r}}^{\mathrm{M}}}^{\mathrm{G2M,features}})$$

$$\mathbf{e}_{v_{\mathrm{s}}^{\mathrm{M}} \to v_{\mathrm{r}}^{\mathrm{G}}}^{\mathrm{M2G}} = \mathrm{MLP}_{\mathcal{E}^{\mathrm{M2G}}}^{\mathrm{embedder}}(\mathbf{e}_{v_{\mathrm{s}}^{\mathrm{M}} \to v_{\mathrm{r}}^{\mathrm{G}}}^{\mathrm{M2G,features}})$$

Prediction pipeline: encoding

$$\mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}}{}' = \mathrm{MLP}^{\mathrm{Grid2Mesh}}_{\mathcal{E}^{\mathrm{G2M}}}([\mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}}, \mathbf{v}^{\mathrm{G}}_{s}, \mathbf{v}^{M}_{r}])$$

$$\mathbf{v}^{\mathrm{M}\prime}_{i} = \mathrm{MLP}^{\mathrm{Grid2Mesh}}_{\mathcal{V}^{\mathrm{M}}}\left(\left[\mathbf{v}^{\mathrm{M}}_{i}, \sum_{e^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}} :\, v^{\mathrm{M}}_{\mathrm{r}}=v^{\mathrm{M}}_{i}} \mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}{}'\right]\right)$$

$$\mathbf{v}^{\mathrm{G}\prime}_{i} = \mathrm{MLP}^{\mathrm{Grid2Mesh}}_{\mathcal{V}^{\mathrm{G}}}\left(\mathbf{v}^{\mathrm{G}}_{i}\right)$$

$$\mathbf{v}^{\mathrm{G}}_{i} \leftarrow \mathbf{v}^{\mathrm{G}}_{i} + \mathbf{v}^{\mathrm{G}\prime}_{i},$$

$$\mathbf{v}^{\mathrm{M}}_{i} \leftarrow \mathbf{v}^{\mathrm{M}}_{i} + \mathbf{v}^{\mathrm{M}\prime}_{i},$$

$$\mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}} \leftarrow \mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}} + \mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}}{}'$$

Prediction pipeline: processing

$$\mathbf{e}^{\mathrm{M}}_{v^M_{\mathrm{s}}\to v^M_{\mathrm{r}}}{}' = \mathrm{MLP}^{\mathrm{Mesh}}_{\mathcal{E}^{\mathrm{M}}}([\mathbf{e}^{\mathrm{M}}_{v^M_{\mathrm{s}}\to v^M_{\mathrm{r}}}, \mathbf{v}^M_s, \mathbf{v}^M_r])$$

$$\mathbf{v}^{\mathrm{M}\prime}_i = \mathrm{MLP}^{\mathrm{Mesh}}_{\mathcal{V}^{\mathrm{M}}}\Big(\Big[\mathbf{v}^{\mathrm{M}}_i, \sum_{e^{\mathrm{M}}_{v^{\mathrm{M}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}:\, v^{\mathrm{M}}_{\mathrm{r}}=v^{\mathrm{M}}_i} \mathbf{e}^{\mathrm{M}}_{v^{\mathrm{M}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}{}'\Big]\Big)$$

$$\mathbf{v}^{\mathrm{M}}_i \leftarrow \mathbf{v}^{\mathrm{M}}_i + \mathbf{v}^{\mathrm{M}\prime}_i$$

$$\mathbf{e}^{\mathrm{M}}_{v^M_{\mathrm{s}}\to v^M_{\mathrm{r}}} \leftarrow \mathbf{e}^{\mathrm{M}}_{v^N_{\mathrm{s}}\to v^M_{\mathrm{r}}} + \mathbf{e}^{\mathrm{M}}_{v^M_{\mathrm{s}}\to v^M_{\mathrm{r}}}{}'$$

Prediction pipeline: decoding

$$\mathbf{e}^{\mathrm{M2G}}_{v^{M}_{\mathrm{s}}\rightarrow v^{\mathrm{G}}_{\mathrm{r}}}{}' = \mathrm{MLP}^{\mathrm{Mesh2Grid}}_{\mathcal{E}^{\mathrm{M2G}}}([\mathbf{e}^{\mathrm{M2G}}_{v^{M}_{\mathrm{s}}\rightarrow v^{\mathrm{G}}_{\mathrm{r}}}, \mathbf{v}^{M}_{s}, \mathbf{v}^{\mathrm{G}}_{r}])$$

$$\mathbf{v}^{\mathrm{G}\prime}_{i} = \mathrm{MLP}^{\mathrm{Mesh2Grid}}_{\mathcal{V}^{\mathrm{G}}}([\mathbf{v}^{\mathrm{G}}_{i}, \sum_{e^{\mathrm{M2G}}_{v^{\mathrm{M}}_{\mathrm{s}}\rightarrow v^{\mathrm{G}}_{\mathrm{r}}} :\, v^{\mathrm{G}}_{\mathrm{r}} = v^{\mathrm{G}}_{i}} \mathbf{e}^{\mathrm{M2G}}_{v^{\mathrm{M}}_{\mathrm{s}}\rightarrow v^{\mathrm{G}}_{\mathrm{r}}}{}'])$$

$$\mathbf{v}^{\mathrm{G}}_{i} \leftarrow \mathbf{v}^{\mathrm{G}}_{i} + \mathbf{v}^{\mathrm{G}\prime}_{i}$$

Prediction pipeline: output

$$\hat{\mathbf{y}}_i^{\mathrm{G}} = \mathrm{MLP}_{\mathcal{V}^{\mathrm{G}}}^{\mathrm{Output}}(\mathbf{v}_i^{\mathrm{G}})$$

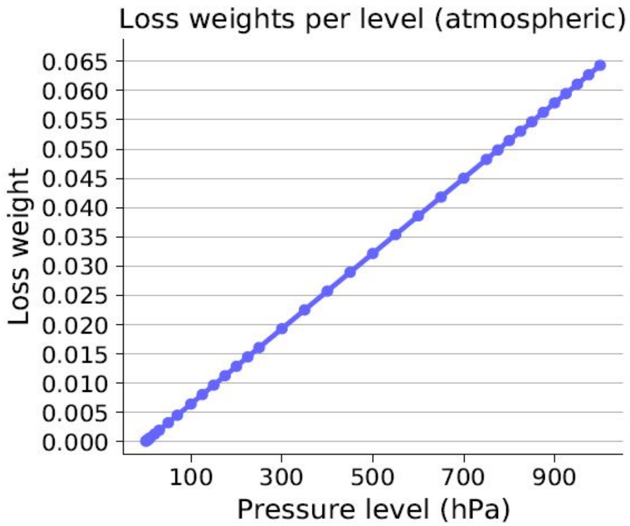
$$\hat{X}^{t+1} = \mathrm{GraphCast}(X^t, X^{t-1}) = X^t + \hat{Y}^t$$

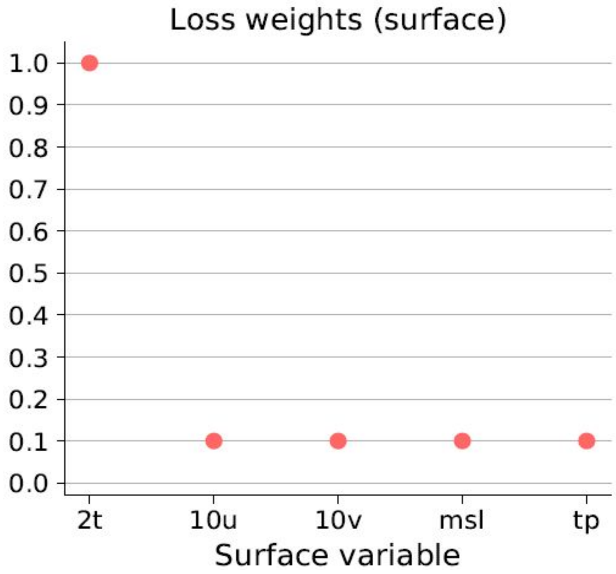
Training

$$\mathcal{L}_{\text{MSE}} = \qquad (\hat{x}_{i,j}^{d_0+\tau} - x_{i,j}^{d_0+\tau})^2$$

Training

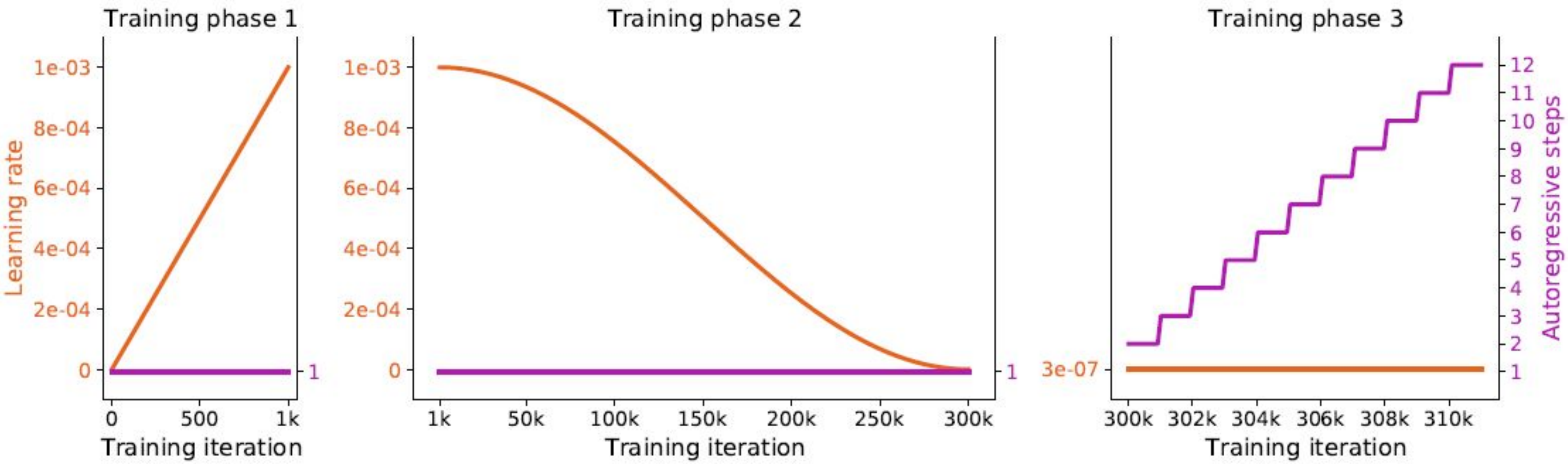
- Data from 1979 to 2017
- Weightings



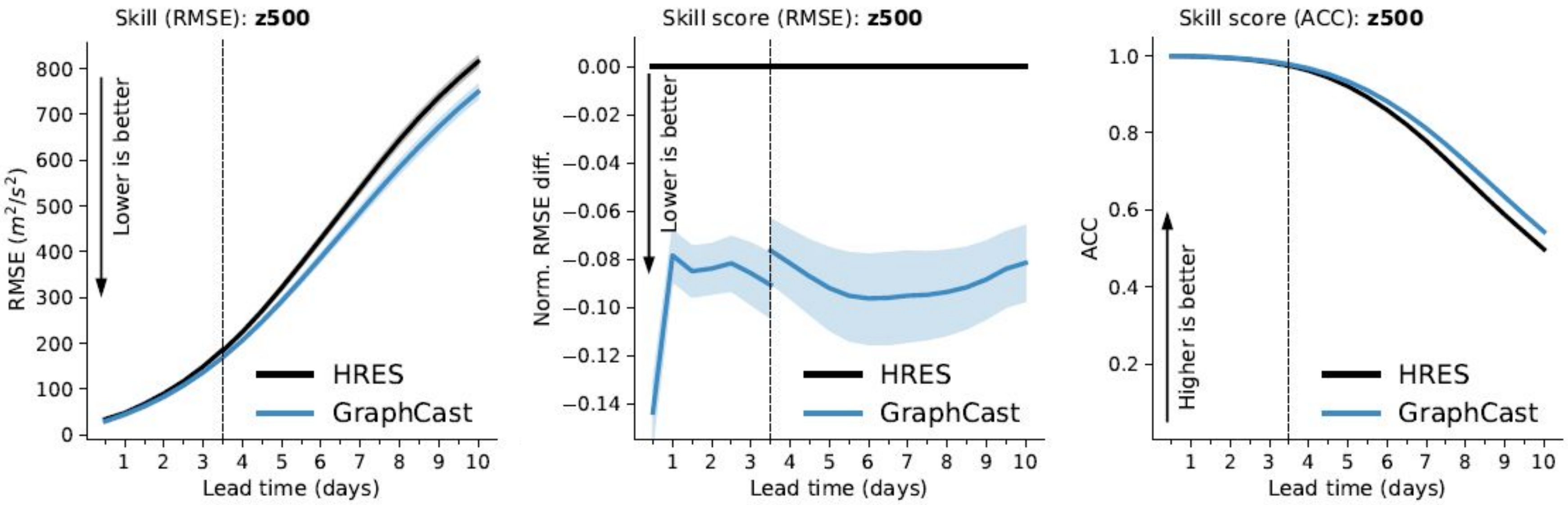


Training

- Data from 1979 to 2017
- Weightings
- Progressively more autoregressive steps

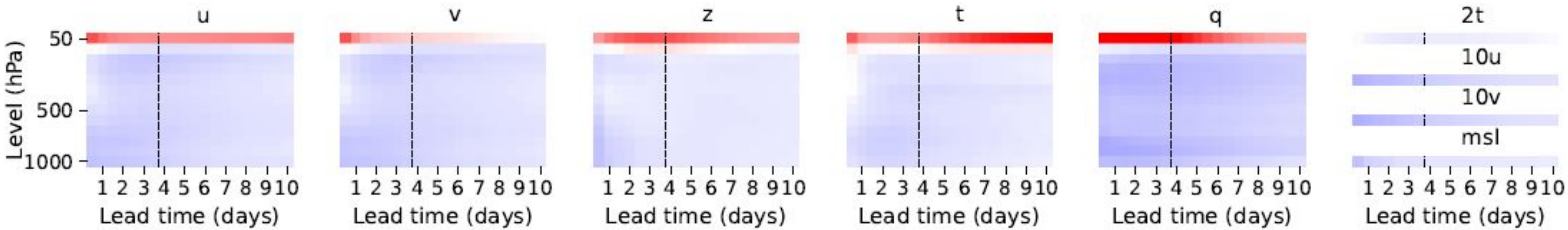


Results

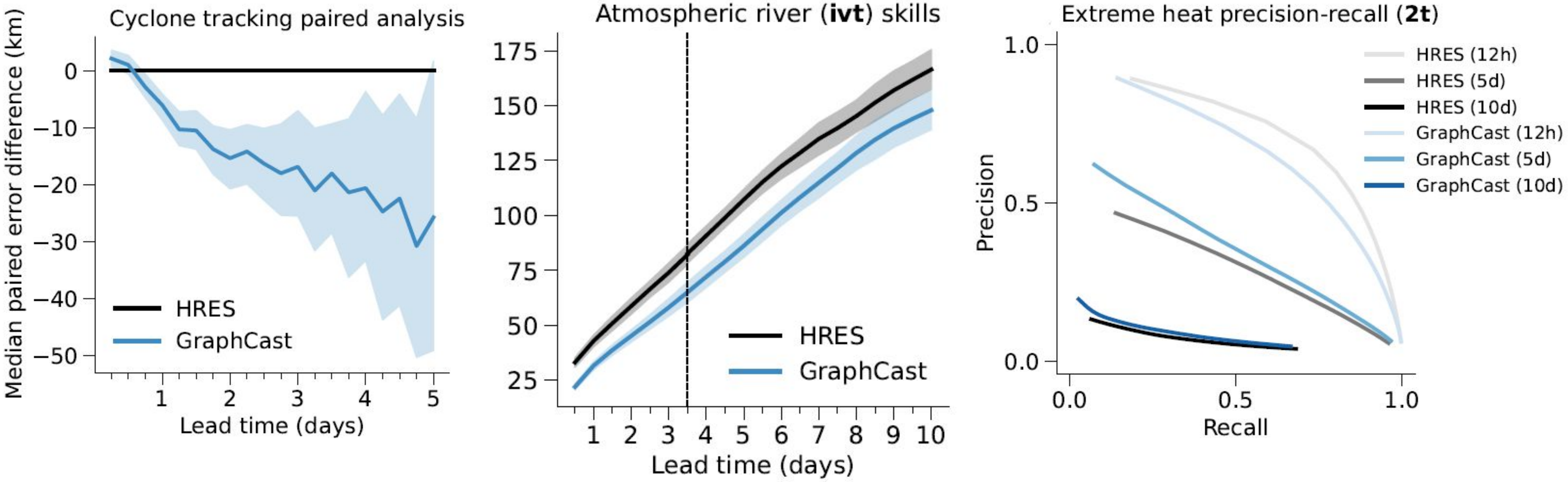


Skill (RMSE): **z500**
RMSE (m^2/s^2)
Lower is better
HRES
GraphCast
Lead time (days)
Skill score (RMSE): **z500**
Norm. RMSE diff.
Lower is better
HRES
GraphCast
Lead time (days)
Skill score (ACC): **z500**
ACC
Higher is better
HRES
GraphCast
Lead time (days)

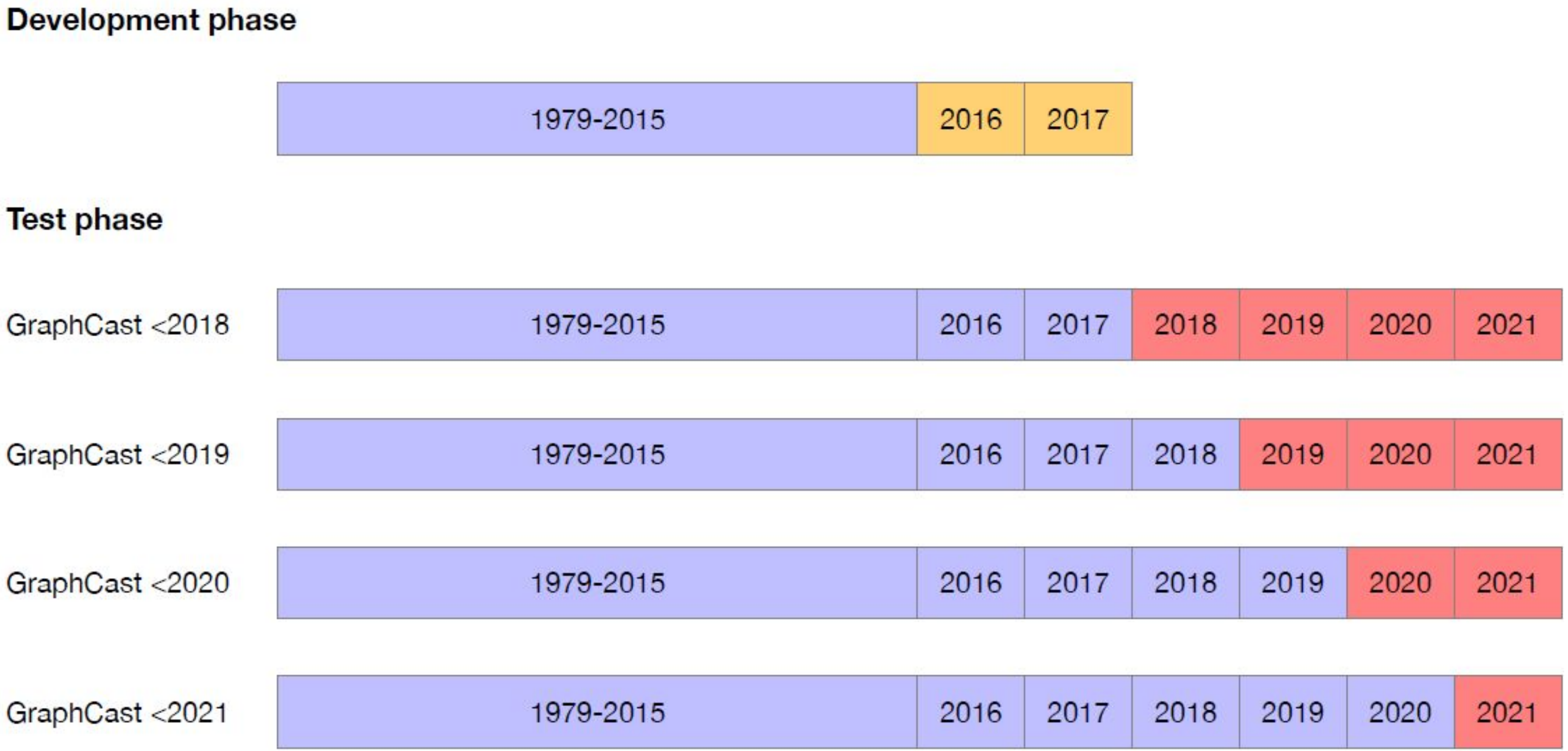
Results


u
v
z
t
q
2t
10u
10v
msl
Level (hPa)
50
500
1000
1 2 3 4 5 6 7 8 9 10
Lead time (days)

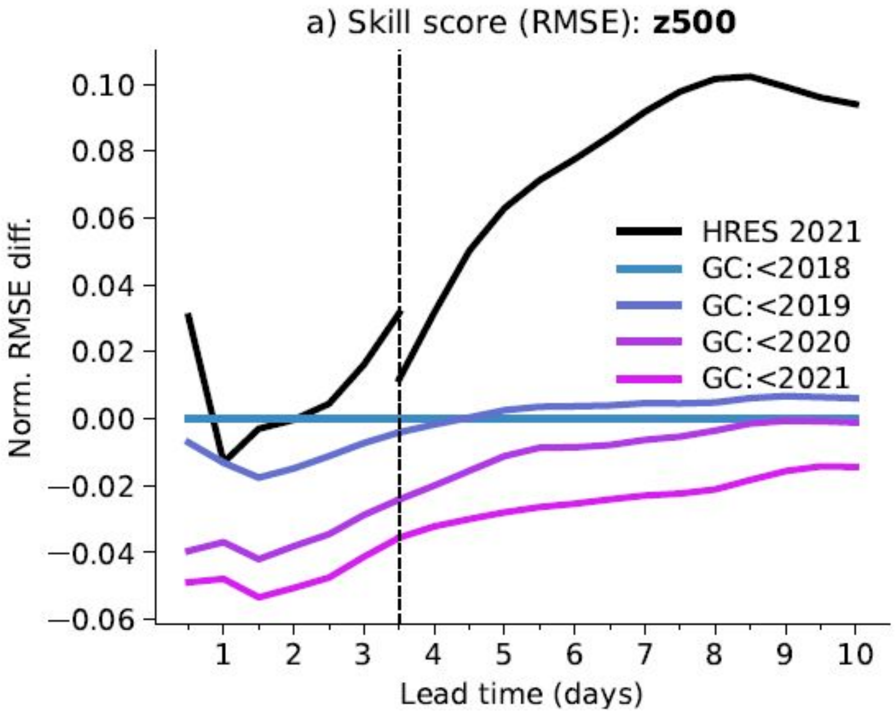
Severe event forecasting



Training data recency



Training data recency



Ablation: mesh

Ablation: autoregressive steps

- Longer horizons blurred more
- Shorter horizons more accurate for short-term forecasting
- A mix of models could be best

Thank you