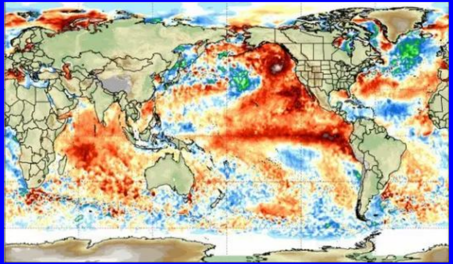


The World as a Graph: Improving El Niño Forecasts with Graph Neural Networks

Salva Rühling Cachay[1], Emma Erickson*[2],
Arthur Fender C. Bucker*[3, 4], Ernest Pokropek*[5], Willa Potosnak*[6],
Suyash Bire[8], Salomey Osei[7], and Björn Lütjens[8]

[1]Technical University of Darmstadt, [2]University of Illinois at Urbana-Champaign,
[3]University of São Paulo, [4] Technical University of Munich, [5]Warsaw University of Technology,
[6]Duquesne University, [7]African Institute for Mathematical Sciences, [8]Massachusetts Institute of Technology

Work motivated by the ***ProjectX*** research competition & supported by a ***Microsoft AI For Earth Grant***

https://arxiv.org/abs/2104.05089

The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

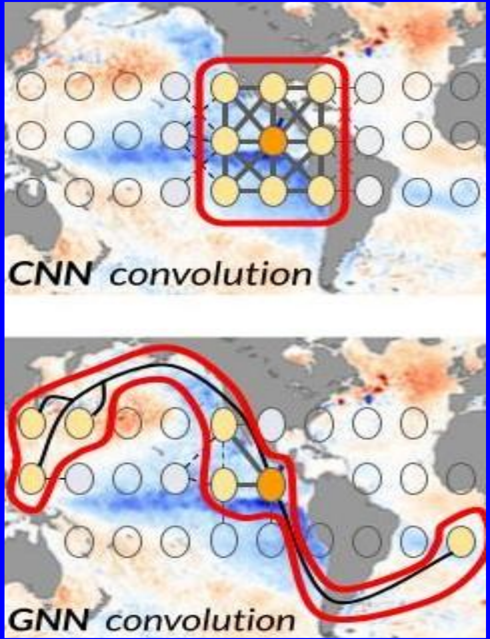
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	**CNNs**:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input

2. Contributions

- **First** application of graph neural networks (**GNN**) to long range forecasting & design of the **Graphiño** GNN.
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**



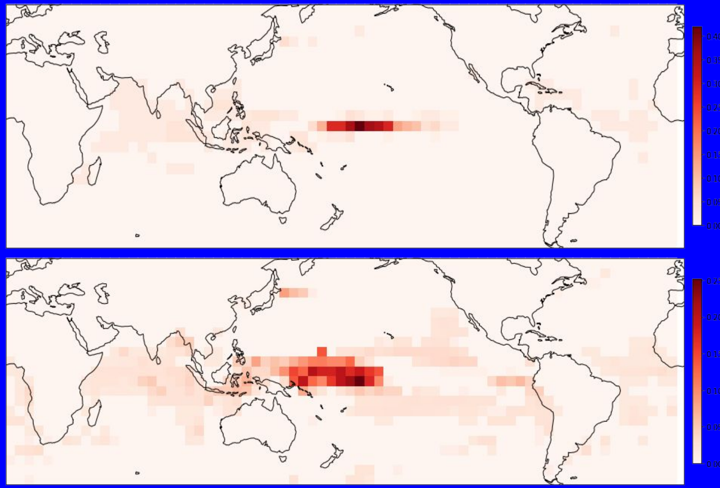


Fig 2. Learned world connectivity structure* for 1 (top) and 6 (bottom) lead months.
***based on the eigenvector centrality score of each node/grid cell**

Model	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=9$	$n=12$	$n=23$
SINTEX-F [37]	0.895	0.89	0.84	0.805	0.78	0.74	0.62	0.51	0.315
CNN [8]	0.9423	0.9158	0.8761	0.8320	0.7983	0.7616	**0.7133**	**0.6515**	0.2870
Graphiño	**0.9747**	**0.9461**	**0.9170**	**0.8742**	**0.8226**	**0.7800**	0.6313	0.5755	**0.3363**

All-season correlation skill in ENSO forecasting for *n lead months, with the same setup as Ham et al. [8]*
Our graph convolutional network+connectivity learning module, ***Graphiño***, is **competitive** to and more **interpretable** (Fig 2.) than previous state-of-the-art, due to its better suited inductive biases.

El Niño–Southern Oscillation (ENSO)

- ENSO warm phase → El Niño
- ENSO cold phase → La Niña
- Causes disasters worldwide
- Mode of climate variability
- ONI (or Niño3.4 index) is a common measure of ENSO
    - Sea surface temperature anomalies in the tropical Pacific (averaged out over the **ONI region** and 3 months)

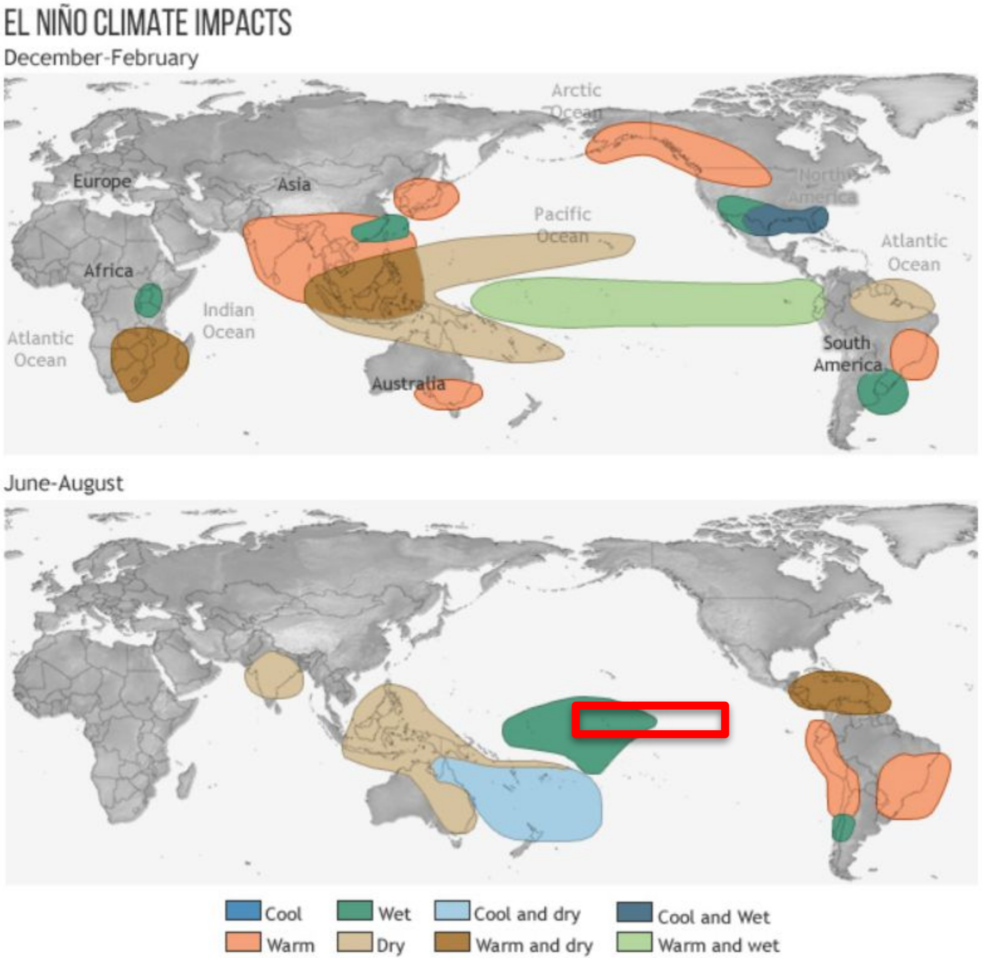


Image from:
https://www.climate.gov/news-features/understanding-climate/2015-state-climate-el-ni%C3%B1o-came-saw-and-conquered

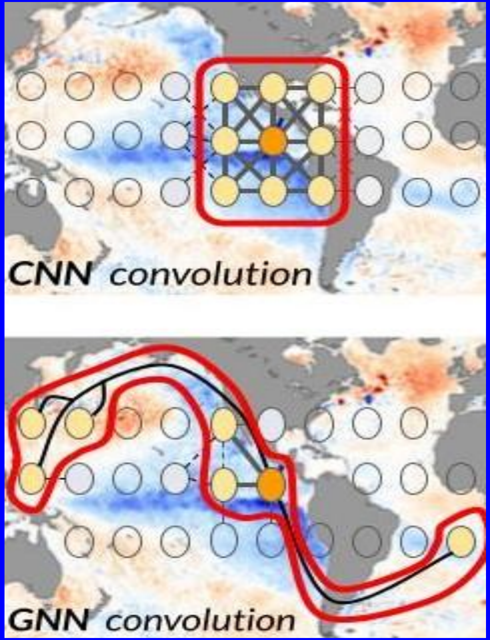
The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

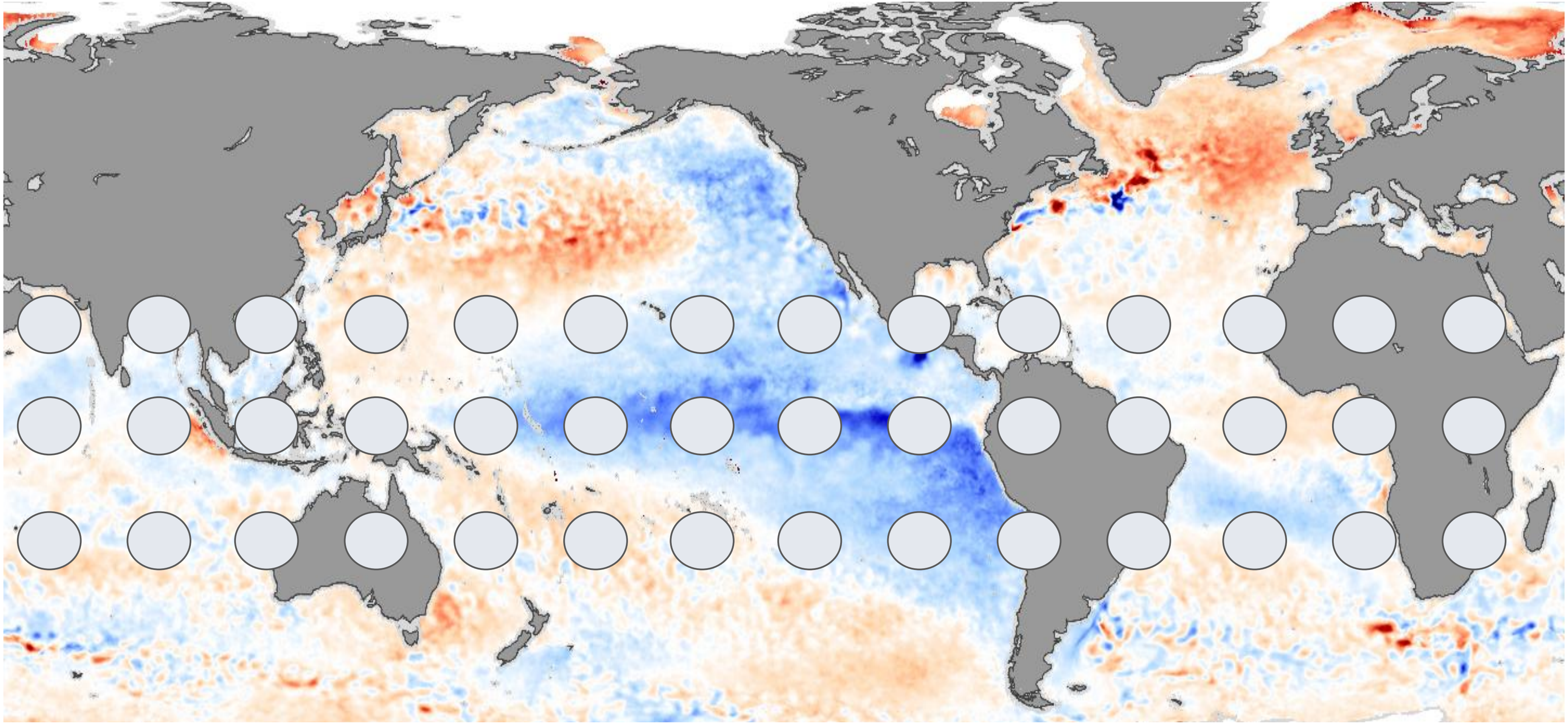
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

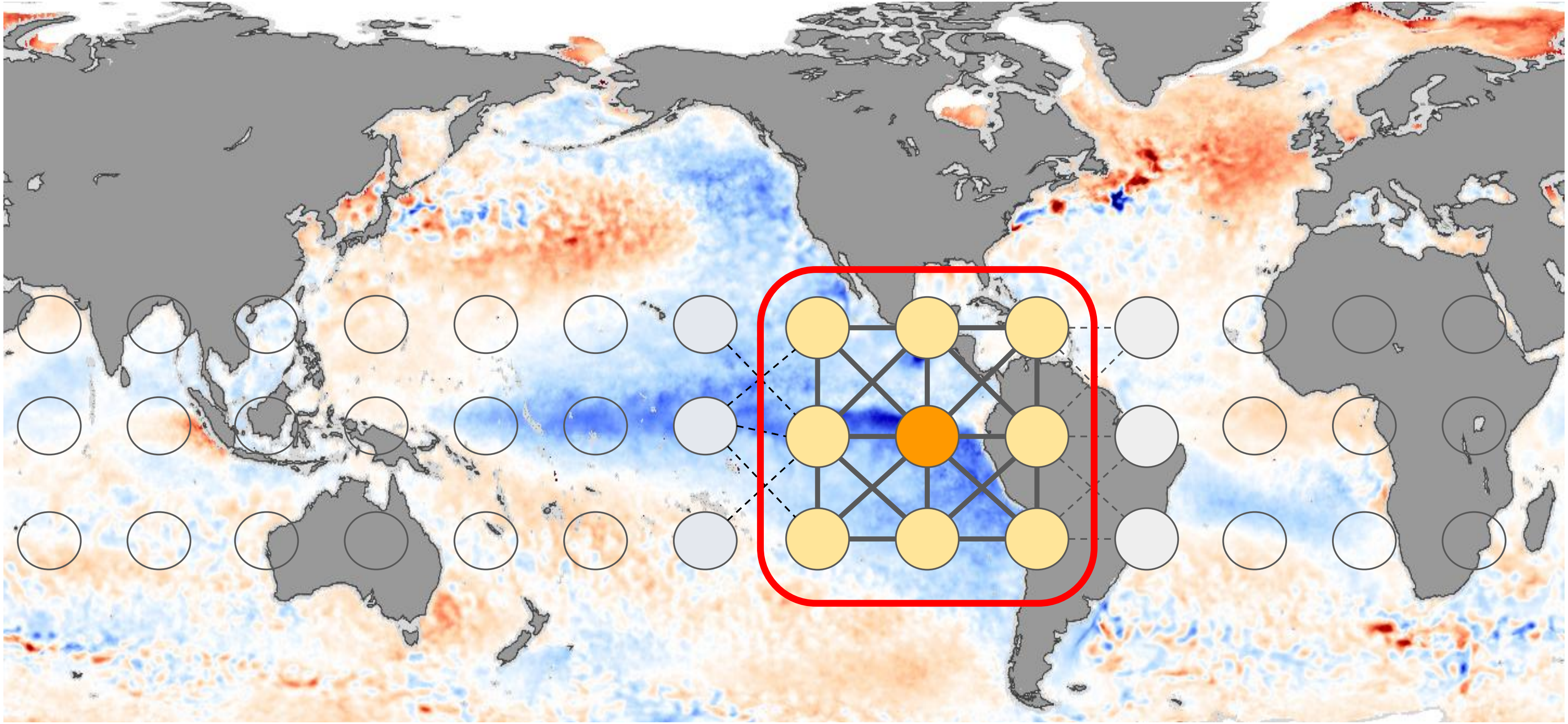
(Long Range) Forecasting:	**CNNs**:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input



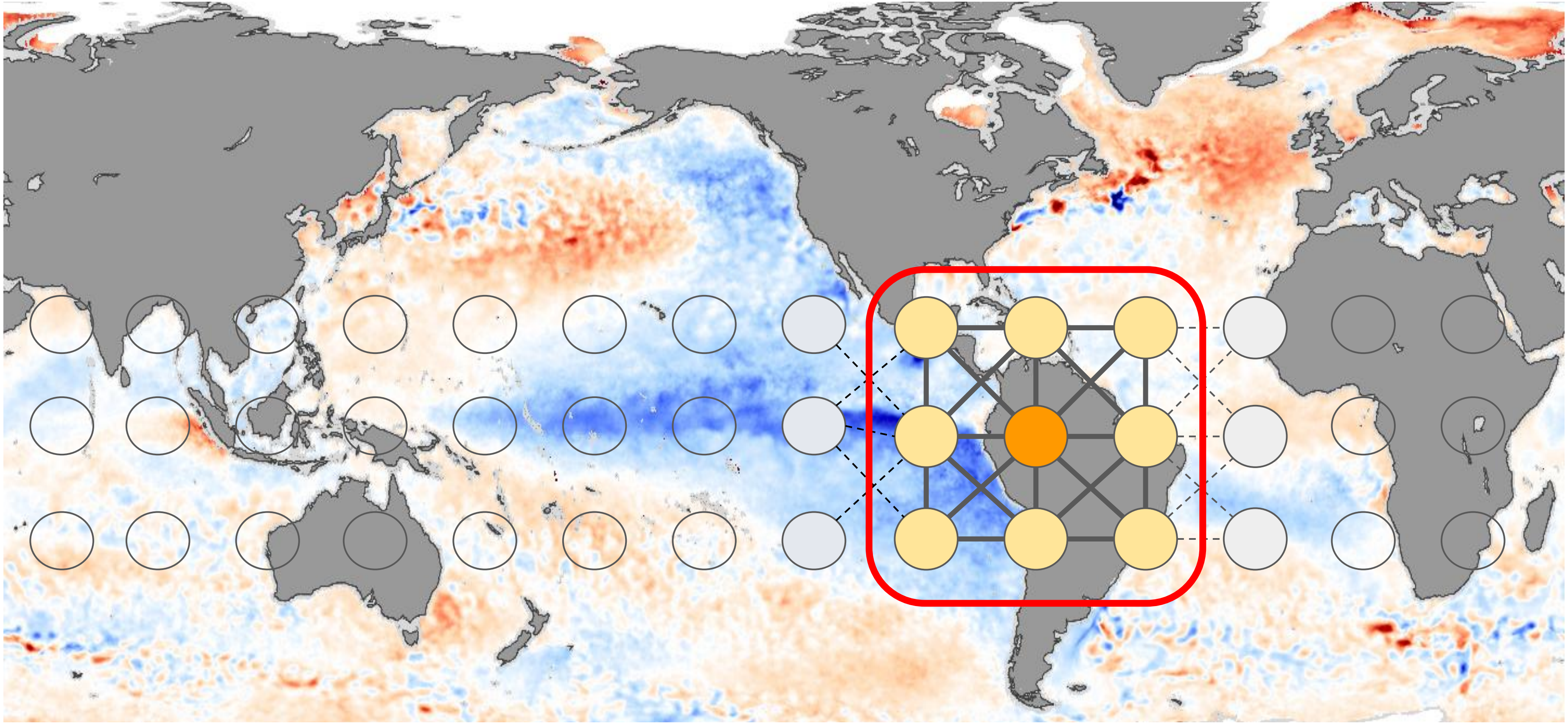
A gridded climate dataset



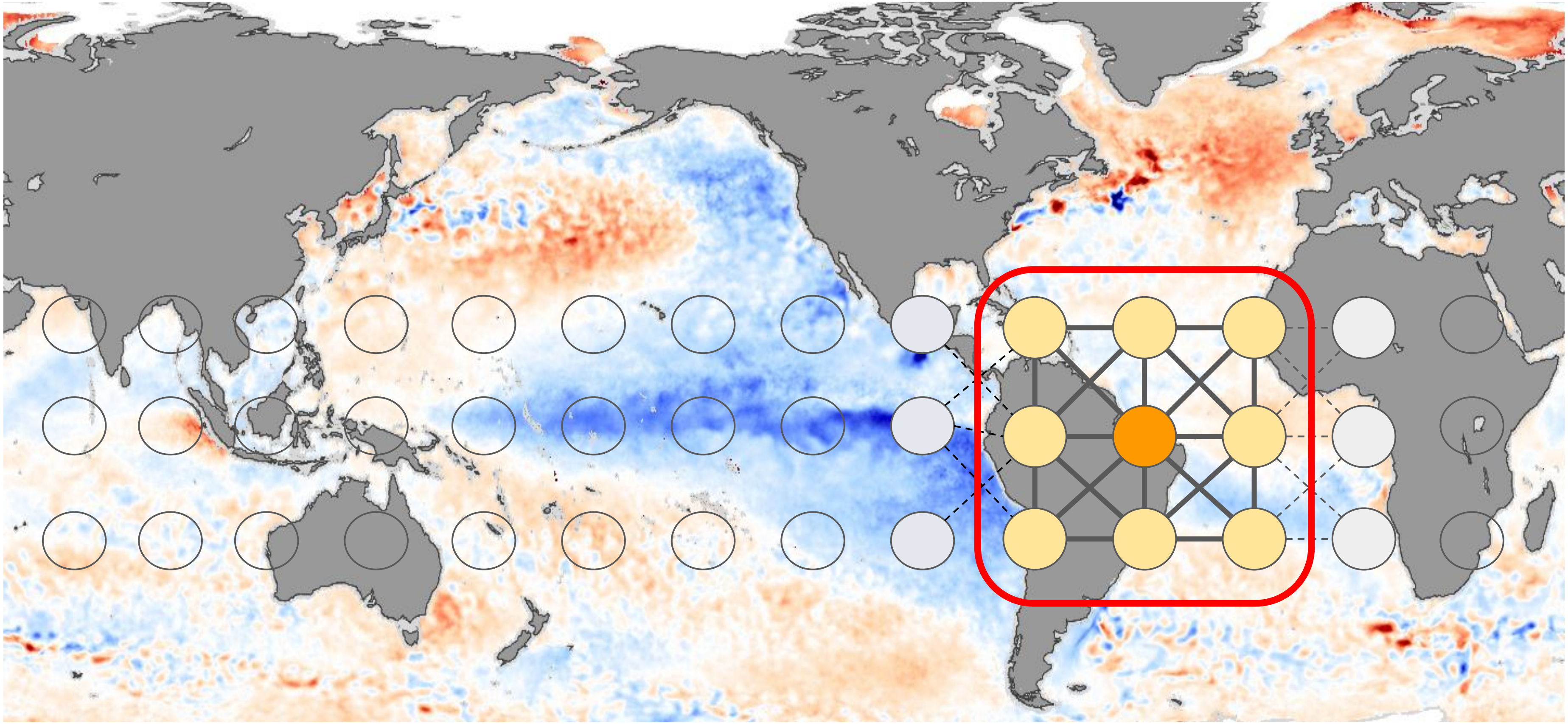
- *Convolutional Neural Network (CNN)* **convolutions** depend on the values of the **center** node and its **<u>local</u>** **neighbors**



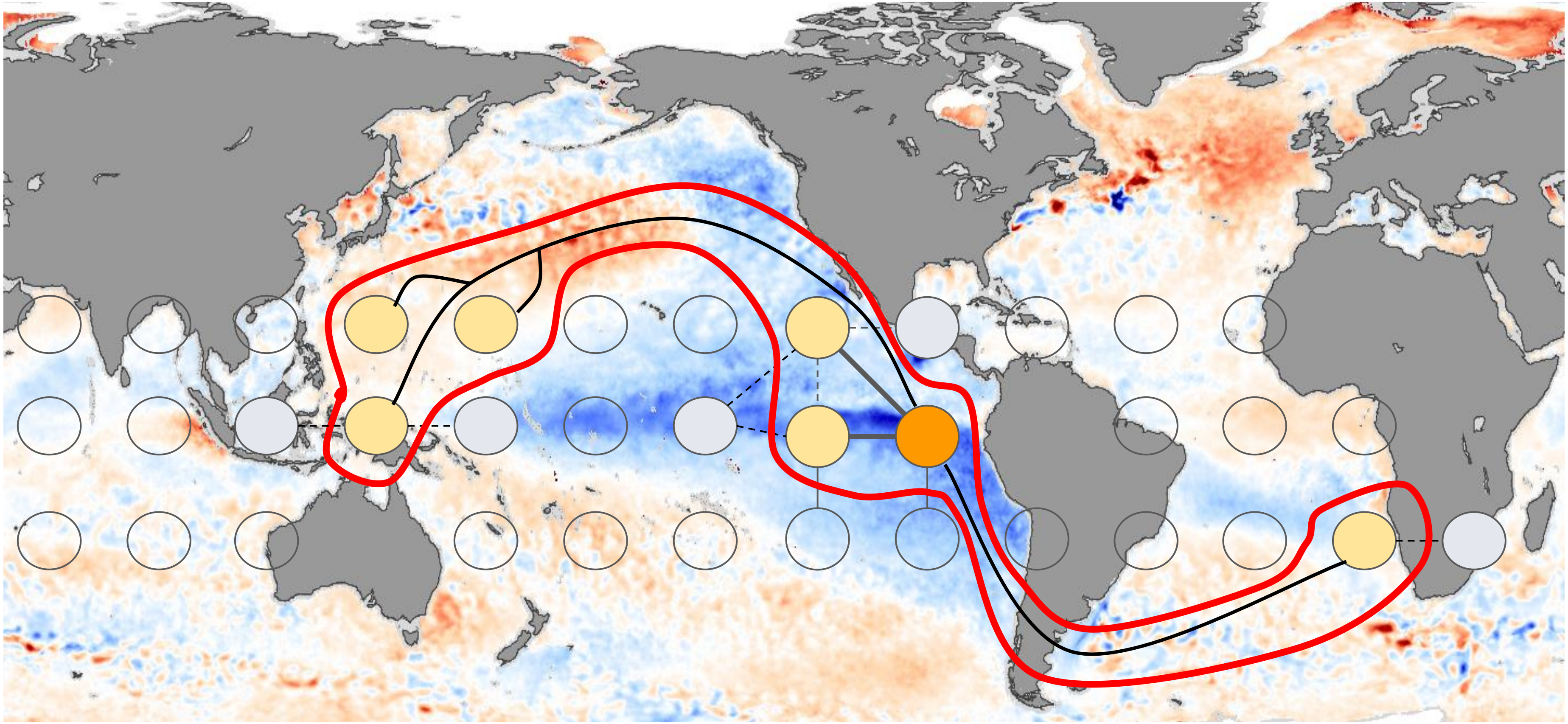
- *Convolutional Neural Network (CNN)* **convolutions** depend on the values of the **center** node and its **local** **neighbors**



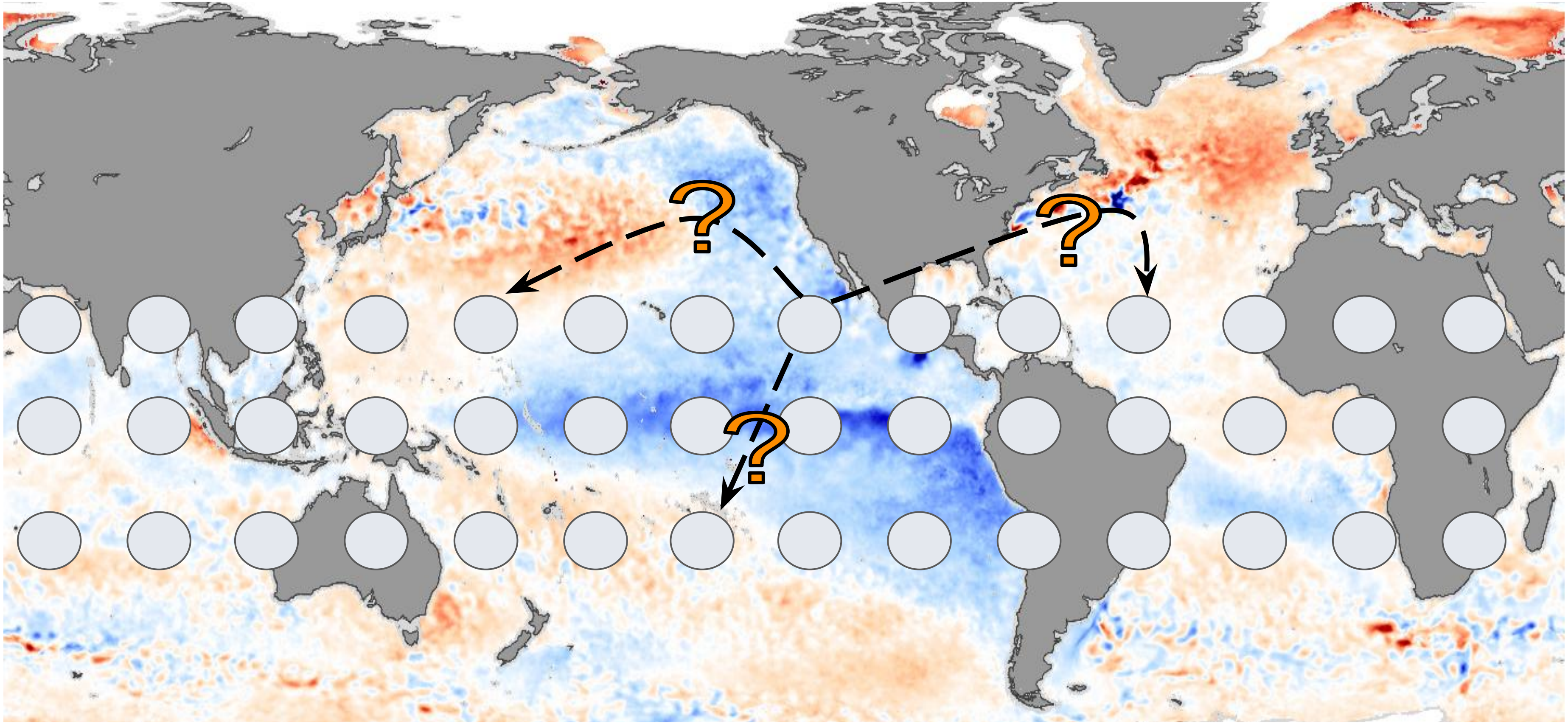
- *Convolutional Neural Network (CNN)* **convolutions** depend on the values of the **center** node and its **<u>local</u>** **neighbors**



– A graph **convolution** *generalizes* them to a variable number of **non-Euclidean** **neighbors** connected by ***edges***



But ..., How do we define the edge structure?



Learn the structure too!

The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

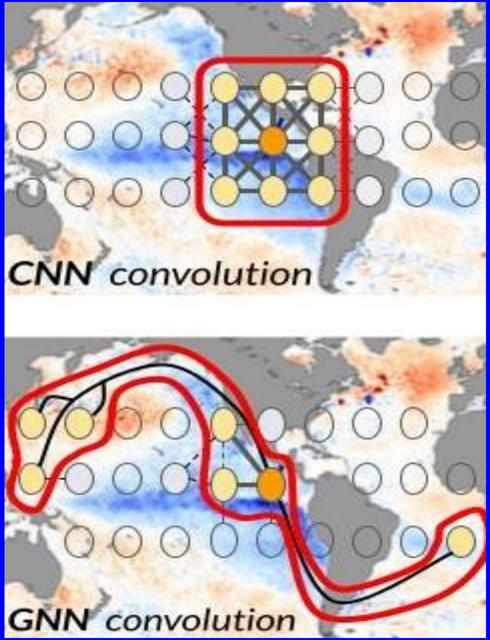
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	**CNNs**:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input

2. Contributions

- **First** application of graph neural networks (**GNN**) to long range forecasting & design of the **Graphiño** GNN
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**



The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

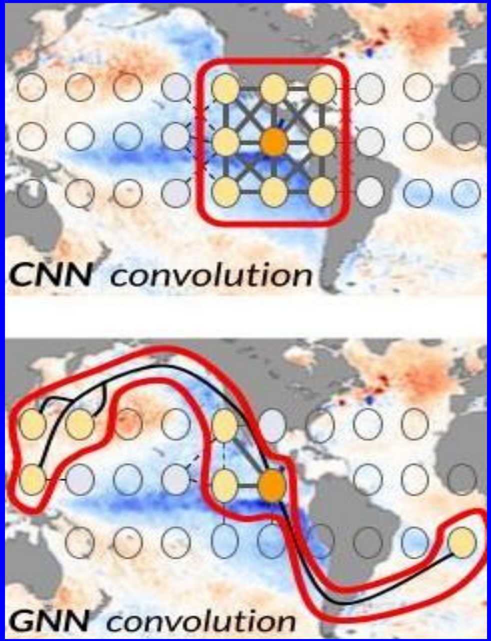
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	**CNNs:**
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input

2. Contributions

- **First** application of graph neural networks (**GNN**) to long range forecasting & design of the **Graphiño** GNN
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**



Learn the structure too!

What do we want?

- Parameter-efficiency
- (weighted) edges in [0, 1]
- Directed edges?
- Sparsitivity

Structure learner

At each iteration:

$$\mathbf{M}_1 = \tanh\left(\alpha_1 \tilde{\mathbf{X}}\tilde{\mathbf{W}}_1\right) \in \mathbb{R}^{N \times \tilde{d}_2}, \quad (1)$$

$$\mathbf{M}_2 = \tanh\left(\alpha_1 \tilde{\mathbf{X}}\tilde{\mathbf{W}}_2\right) \in \mathbb{R}^{N \times \tilde{d}_2}, \quad (2)$$

$$\mathbf{A} = \mathrm{sigmoid}\left(\alpha_2 \mathbf{M}_1 \mathbf{M}_2^T\right) \in [0,1]^{N \times N}, \quad (3)$$

where $\tilde{\mathbf{W}}_1, \tilde{\mathbf{W}}_2 \in \mathbb{R}^{\tilde{d}_1 \times \tilde{d}_2}$ are learnable parameters,

(4) Remove all but the largest E values A_ij (set them to zero)

- Parameter-efficiency

$N-$	The number of nodes
$\tilde{\mathbf{X}} \in \mathbb{R}^{N \times \tilde{d}_1} -$	Static node representations

Builds upon & is Inspired by: Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "*Connecting the dots: Multivariate time series forecasting with graph neural networks,*" KDD 2020

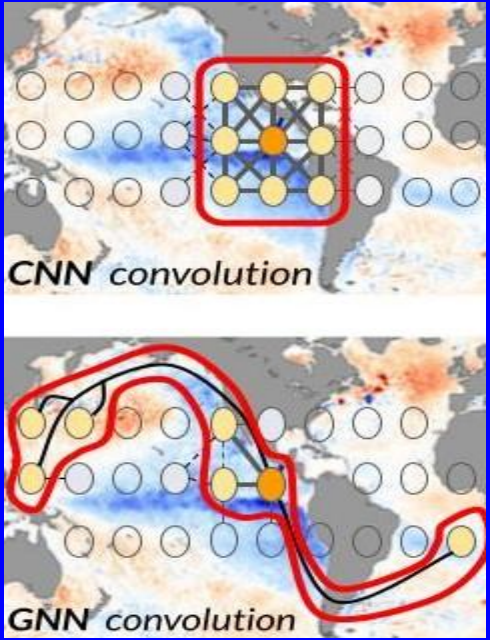
The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

1. Motivation

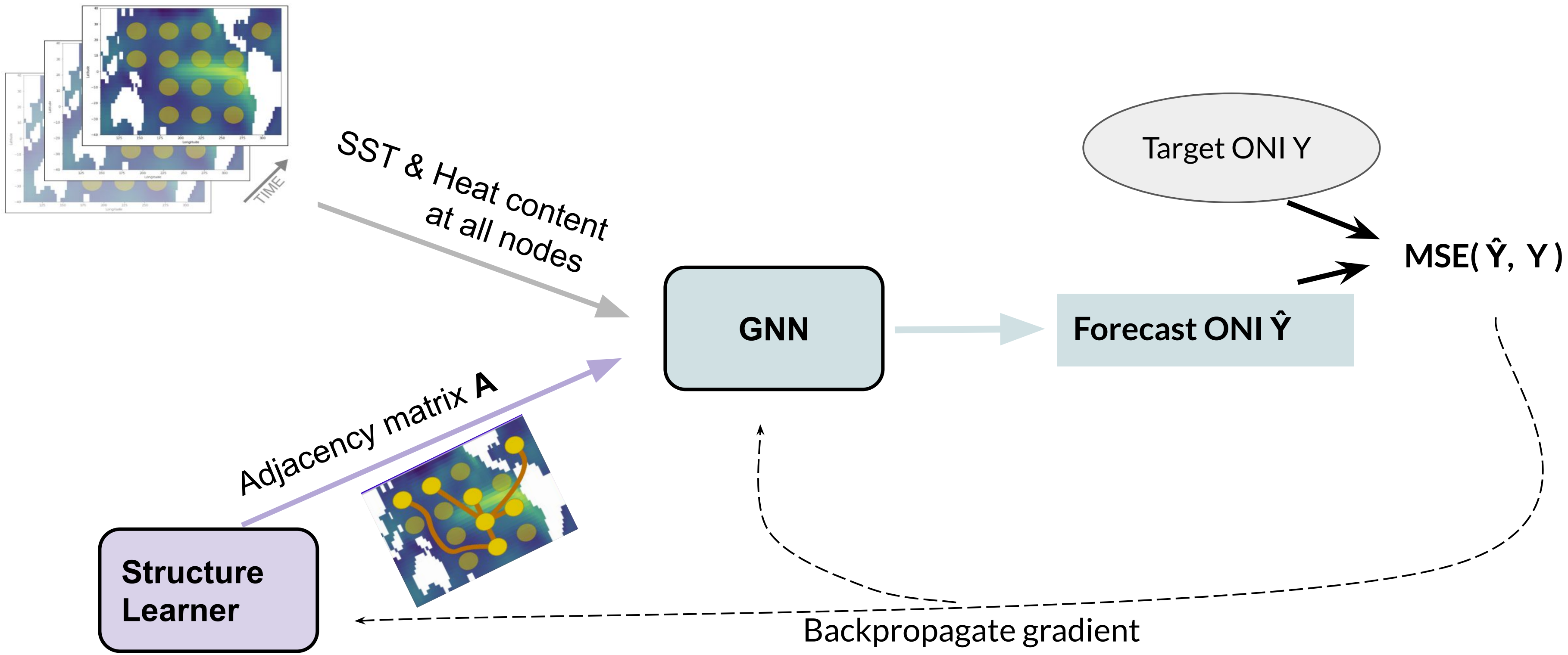
Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	CNNs:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input



2. Contributions

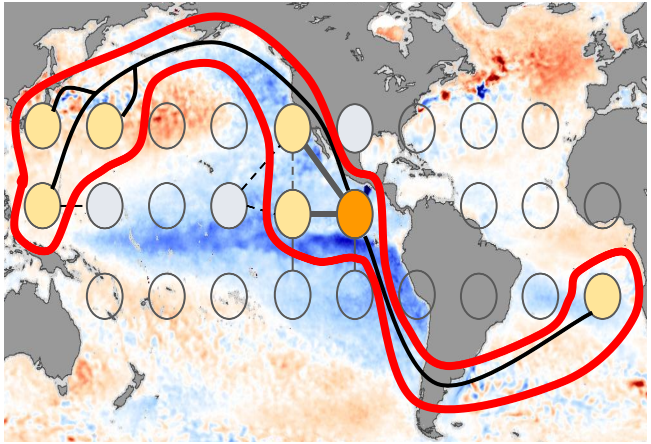
- **First** application of graph neural networks (**GNN**) to long range forecasting & **design of the Graphiño GNN**
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**


TIME
SST & Heat content
at all nodes
Target ONI Y
MSE(Ŷ, Y)
GNN
Forecast ONI Ŷ
Adjacency matrix **A**
Structure Learner
Backpropagate gradient

The Graph Neural Network

- Projecting the ONI can be framed as a *graph regression* problem
- We build upon a *Graph Convolutional Network* (GCN; Kipf et al.)
- A basic **graph convolution** can be written as

$$\mathbf{Z} = \sigma(\mathbf{AXW}) \in \mathbb{R}^{N \times \text{out-dim}}$$



- The GCN is extended by
  - *residual* and *jumping knowledge* connections
  - Node-in-degree normalization replaced with *batch-normalization over feature dimension*

The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

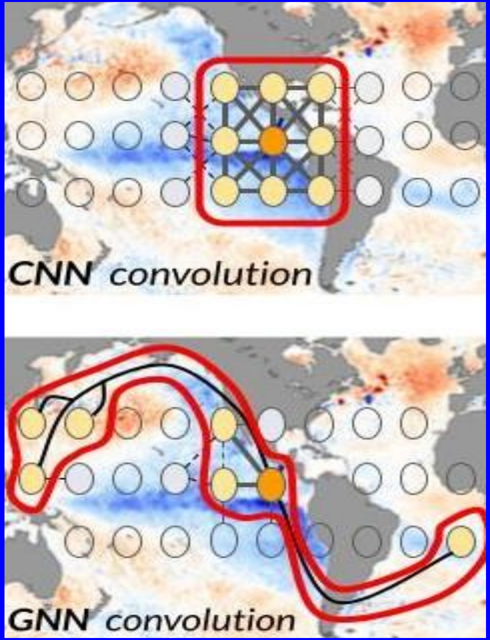
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	CNNs:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input

2. Contributions

- **First** application of graph neural networks (**GNN**) to long range forecasting & design of the **Graphiño** GNN
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**



Model	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=9$	$n=12$	$n=23$
SINTEX-F [37]	0.895	0.89	0.84	0.805	0.78	0.74	0.62	0.51	0.315
CNN [8]	0.9423	0.9158	0.8761	0.8320	0.7983	0.7616	**0.7133**	**0.6515**	0.2870
Graphiño	**0.9747**	**0.9461**	**0.9170**	**0.8742**	**0.8226**	**0.7800**	0.6313	0.5755	**0.3363**

All-season correlation skill in ENSO forecasting for *n lead months, with the same setup as Ham et al. [8]*
Our graph convolutional network+connectivity learning module, ***Graphiño***, is **competitive** to and more **interpretable** (Fig 2.) than previous state-of-the-art, due to its better suited inductive biases.

The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

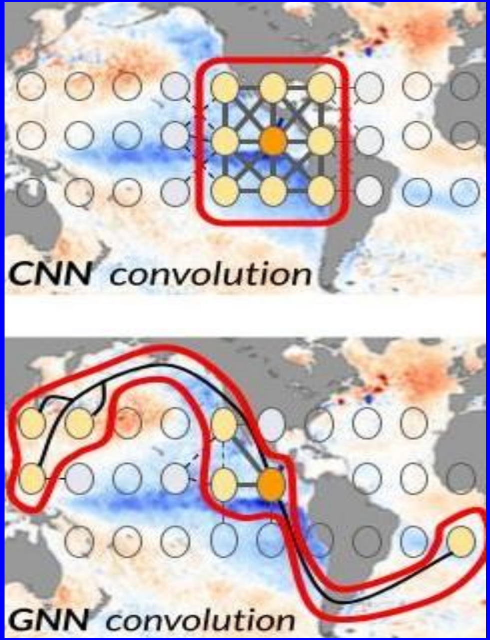
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	**CNNs**:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input

2. Contributions

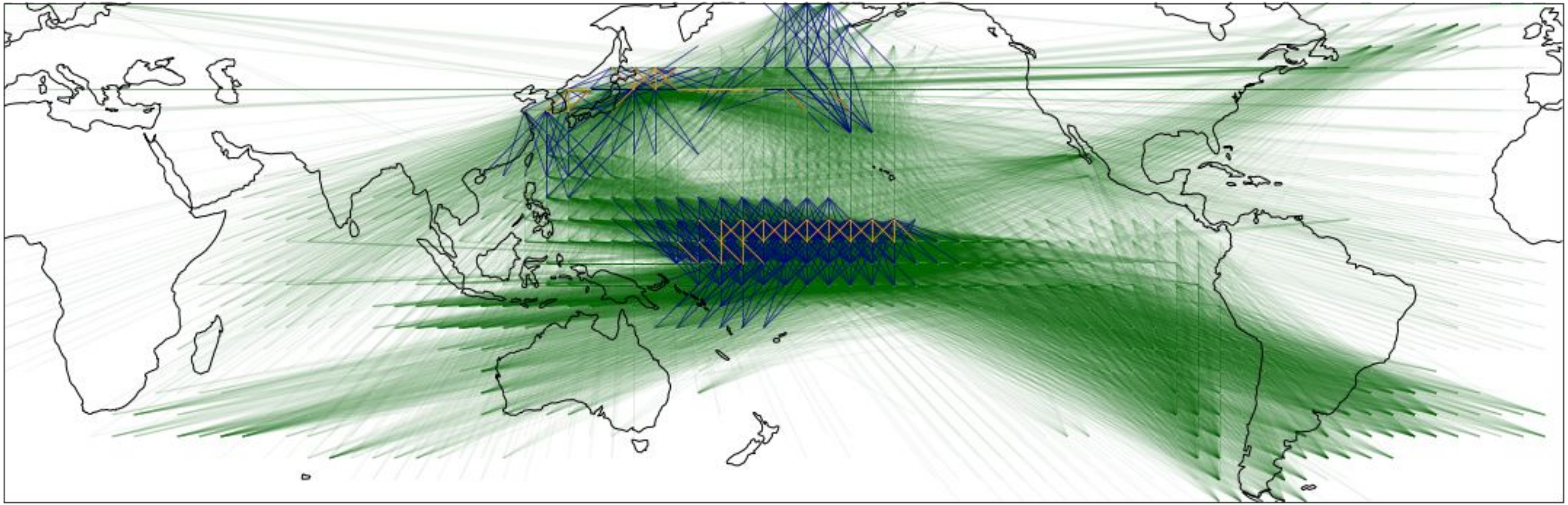
- **First** application of graph neural networks (**GNN**) to long range forecasting & design of the **Graphiño** GNN
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**



Model	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=9$	$n=12$	$n=23$
SINTEX-F [37]	0.895	0.89	0.84	0.805	0.78	0.74	0.62	0.51	0.315
CNN [8]	0.9423	0.9158	0.8761	0.8320	0.7983	0.7616	**0.7133**	**0.6515**	0.2870
Graphiño	**0.9747**	**0.9461**	**0.9170**	**0.8742**	**0.8226**	**0.7800**	0.6313	0.5755	**0.3363**

All-season correlation skill in ENSO forecasting for *n lead months, with the same setup as Ham et al. [8]*
Our graph convolutional network+connectivity learning module, ***Graphiño***, is **competitive** to and more **interpretable** (Fig 2.) than previous state-of-the-art, due to its better suited inductive biases.

How to analyze the learned connectivity?



Around 12k edges/connections...

Eigenvector centrality...

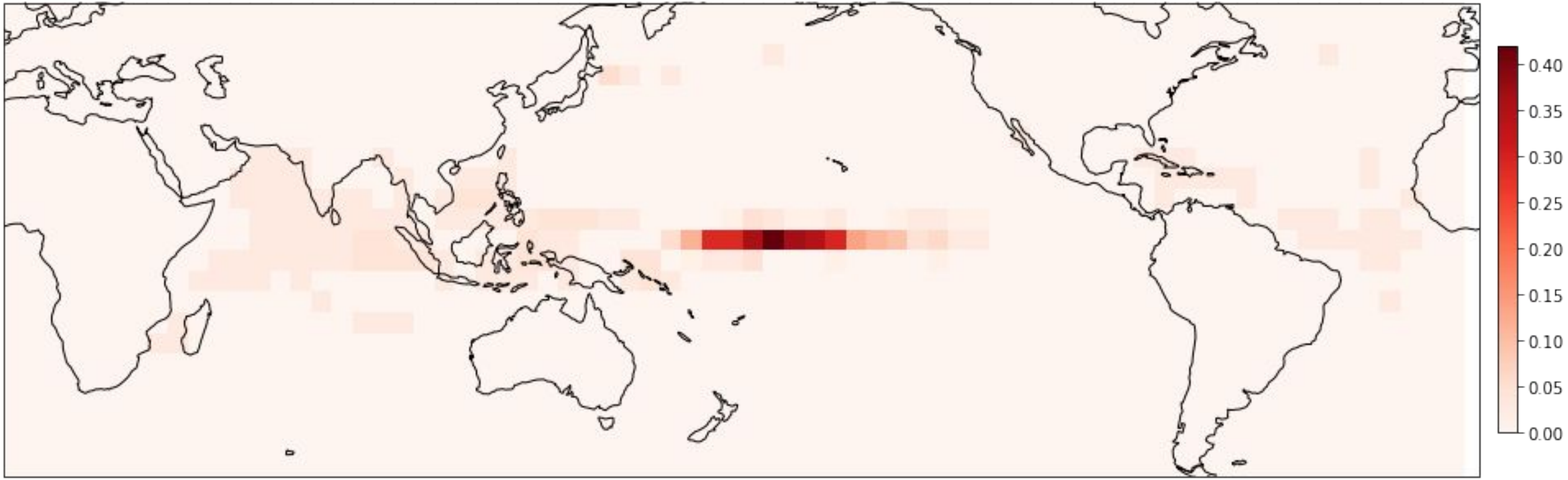
- Measures the **importance**/**influence** of a node in/on the graph

- Google's early Pagerank algorithm is a variant of it
  - Pages with more links (from other important pages) are more important
  - → **Locations with more connections are more important**

    (propagate more information during message-passing/graph convolutions)
- Node centrality vector **v** solves the eigenvector equation of the adjacency matrix: $\mathbf{A}\lambda = \mathbf{v}\lambda$

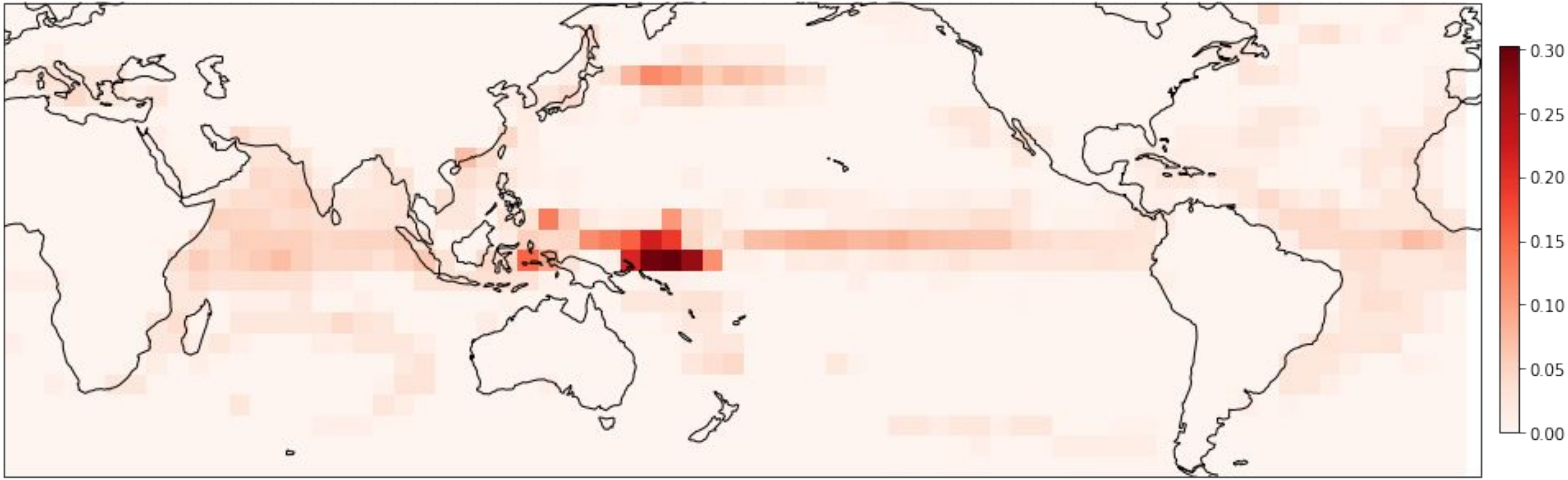
Inspecting the learned world connectivity

→ **1 lead month**



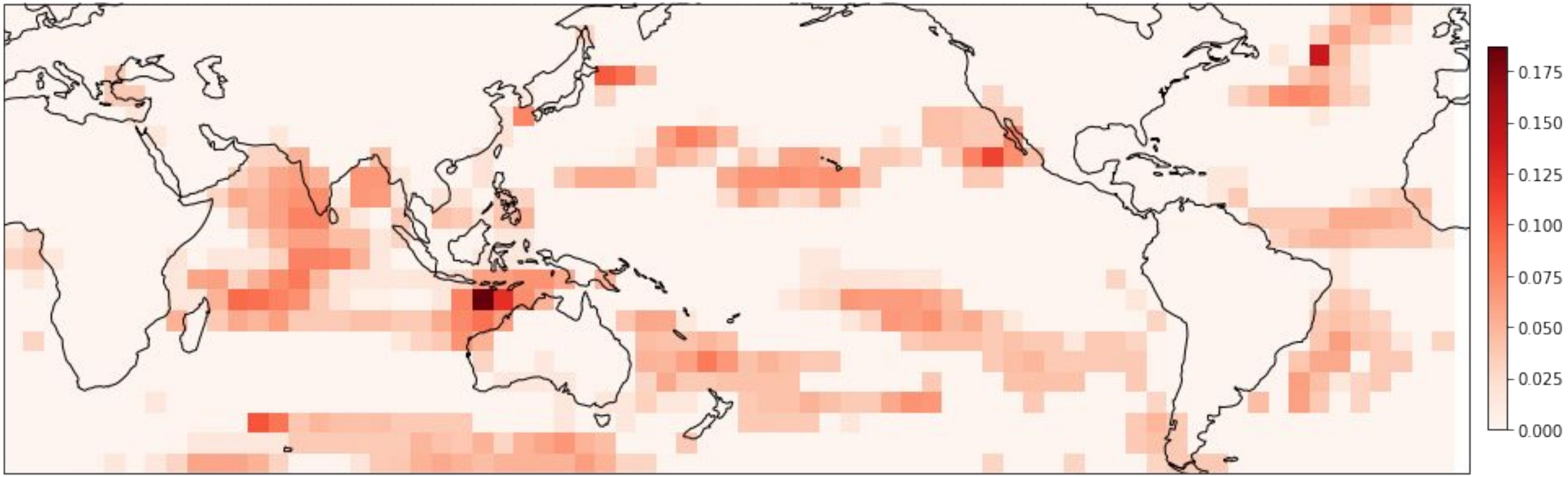
Inspecting the learned world connectivity

→ 3 lead months



Inspecting the learned world connectivity

→ **23 lead months**



Conclusions

- Our proposed structure learning+GNN model **outperforms** competitive dynamical and deep learning models for **up to 6 months**.
- Novel ML **interpretability** method for the earth sciences
- Easily **applicable to related problems** in long range forecasting and beyond

Exciting Future Research Directions...

- GNNs likely not the end of the story...
- Better skill in forecasting **extreme** ENSO **events** (e.g. via a custom loss function)
- **Better structure learning modules** & analysis to potentially find yet undiscovered sources of predictability for ENSO
- Injecting climatologists' **domain knowledge** into pre-defined, fixed connectivity structures

The World as a Graph: Improving El Niño Forecasts with Graph Neural Nets

Salva Rühling Cachay, Emma Erickson, Arthur Fender C. Bucker, Ernest Pokropek, Willa Potosnak, Suyash Bire, Salomey Osei, and Björn Lütjens
salvaruehling@gmail.com

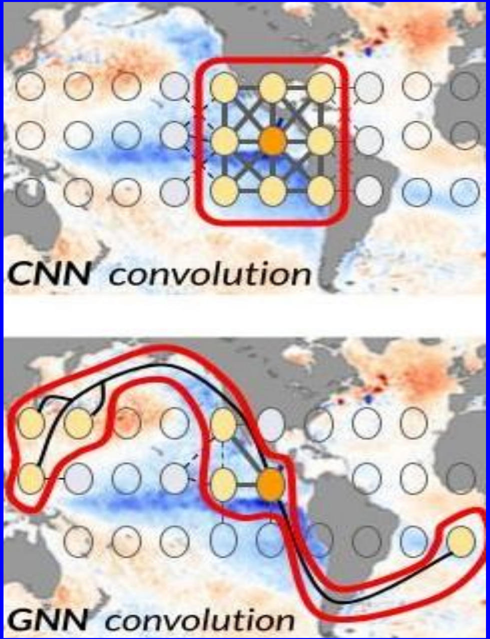
1. Motivation

Deep learning successes in long range forecasting rely on convolutional neural networks (CNN), **but...**

(Long Range) Forecasting:	**CNNs**:
Driven by large-scale/***global*** interactions	Based on spatially ***local*** computations/convolutions
Anomalies/Patterns in different parts of the world should be treated differently	Assume ***translational equivariance***
May only need parts of the world as input	Require a ***grid*** as input

2. Contributions

- **First** application of graph neural networks (**GNN**) to long range forecasting & design of the **Graphiño** GNN.
- Design of a **novel graph connectivity** learning module, → our approach applicable *without a pre-defined graph*
- Model **outperforms** competitive dynamical and deep learning model for *up to 6 months*
- Model **learns a sensible world connectivity structure** that can be **linked to ENSO dynamics theories**



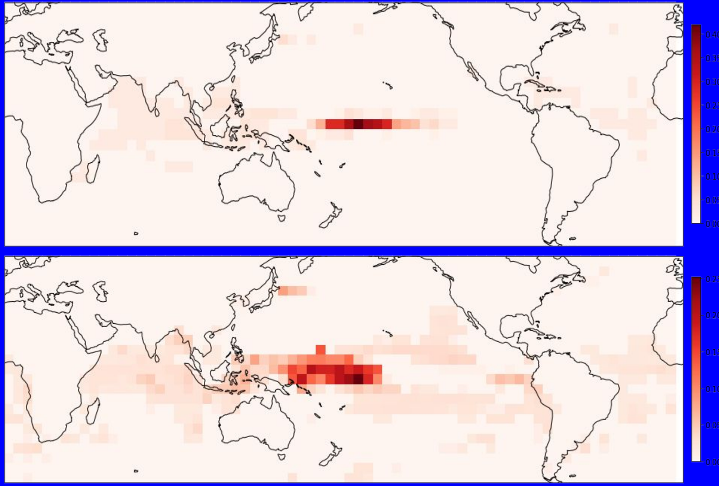
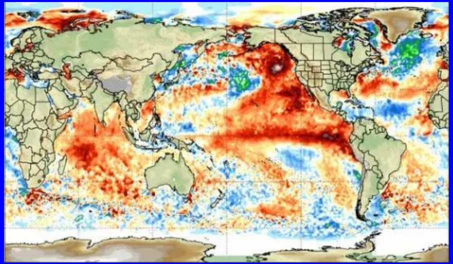


Fig 2. Learned world connectivity structure* for 1 (top) and 6 (bottom) lead months.
***based on the eigenvector centrality score of each node/grid cell**

Model	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 9$	$n = 12$	$n = 23$
SINTEX-F [37]	0.895	0.89	0.84	0.805	0.78	0.74	0.62	0.51	0.315
CNN [8]	0.9423	0.9158	0.8761	0.8320	0.7983	0.7616	**0.7133**	**0.6515**	0.2870
Graphiño	**0.9747**	**0.9461**	**0.9170**	**0.8742**	**0.8226**	**0.7800**	0.6313	0.5755	**0.3363**

All-season correlation skill in ENSO forecasting for *n lead months, with the same setup as Ham et al. [8]*
Our graph convolutional network+connectivity learning module, ***Graphiño***, is **competitive** to and more **interpretable** (Fig 2.) than previous state-of-the-art, due to its better suited inductive biases.



The World as a Graph: Improving El Niño Forecasts with Graph Neural Networks

Salva Rühling Cachay[1], Emma Erickson[2],
Arthur Fender C. Bucker[3, 4], Ernest Pokropek[5], Willa Potosnak[6],
Salomey Osei[7], and Björn Lütjens[8]

[1]Technical University of Darmstadt, [2]University of Illinois at Urbana-Champaign,
[3]University of São Paulo, [4] Technical University of Munich, [5]Warsaw University of Technology,
[6]Duquesne University, [7]African Institute for Mathematical Sciences, [8] Massachusetts Institute of Technology

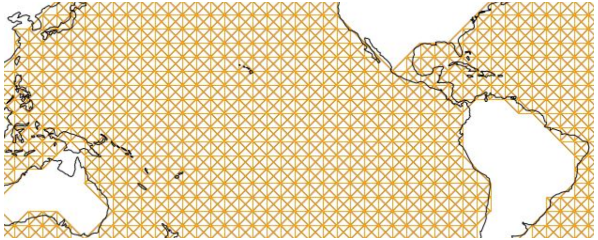
Thank you! Questions?

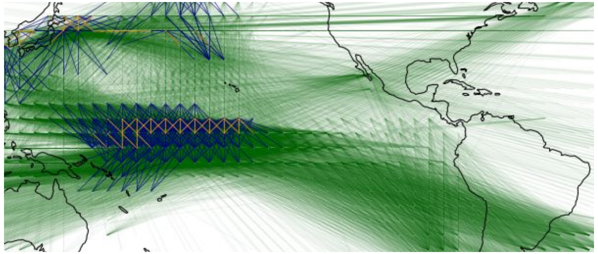
Work motivated by the ***ProjectX*** research competition & supported by a ***Microsoft AI For Earth Grant***

A non-local graph structure is key

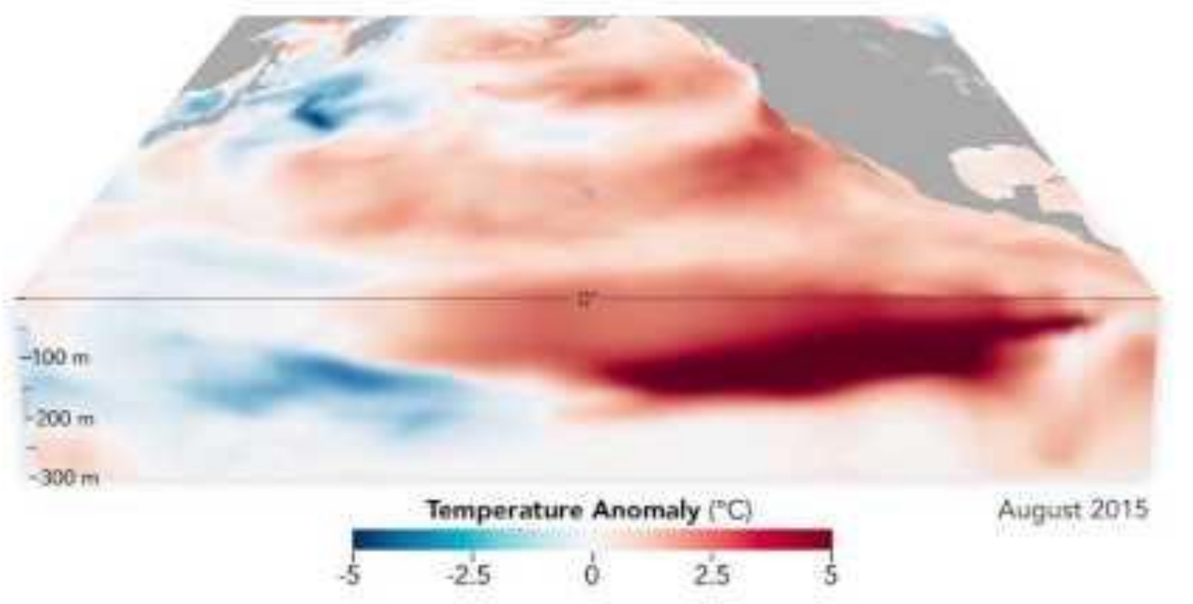
TABLE II: Incorporating geographical distant information is key for a strong performance. We report the correlation skill for n lead months of the same GNN with 1) our structure learning module, 2) the structure learning module from [33], and 3) a fixed, local connectivity structure with edges based on spatial proximity (local).

Edge structure	$n = 1$	$n = 3$	$n = 6$	$n = 9$
Local	0.9063	0.7752	0.5946	0.4586
[33]	0.9117	0.8503	0.6439	0.4190
Graphiño	**0.9747**	**0.9170**	**0.7800**	**0.6313**





[33] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, “*Connecting the dots: Multivariate time series forecasting with graph neural networks,*” KDD 2020



https://earthobservatory.nasa.gov/features/ElNino
https://www.youtube.com/watch?v=69N494UIIS8&t=2s

Structure learner

At each iteration:

$$\mathbf{M}_1 = \tanh\left(\alpha_1 \tilde{\mathbf{X}}\tilde{\mathbf{W}}_1\right) \in \mathbb{R}^{N\times \tilde{d}_2}, \tag{1}$$

$$\mathbf{M}_2 = \tanh\left(\alpha_1 \tilde{\mathbf{X}}\tilde{\mathbf{W}}_2\right) \in \mathbb{R}^{N\times \tilde{d}_2}, \tag{2}$$

$$\mathbf{A} = \mathrm{sigmoid}\left(\alpha_2 \mathbf{M}_1\mathbf{M}_2^T\right) \in [0,1]^{N\times N}, \tag{3}$$

where $\tilde{\mathbf{W}}_1, \tilde{\mathbf{W}}_2 \in \mathbb{R}^{\tilde{d}_1\times \tilde{d}_2}$ are learnable parameters,

(4) Remove all but the largest E values A_ij (set them to zero)

Builds upon & is Inspired by: Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "*Connecting the dots: Multivariate time series forecasting with graph neural networks,*" KDD 2020