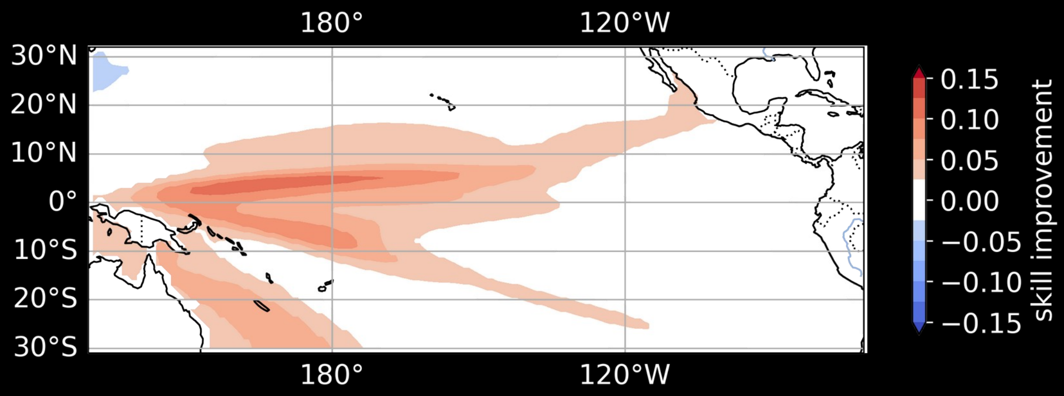
Characterizing Nonlinearities in CESM2 ENSO Dynamics using Machine Learning Technique



Jakob Schlör, Antonietta Capotondi, Matthew Newman, Bedartha Goswami

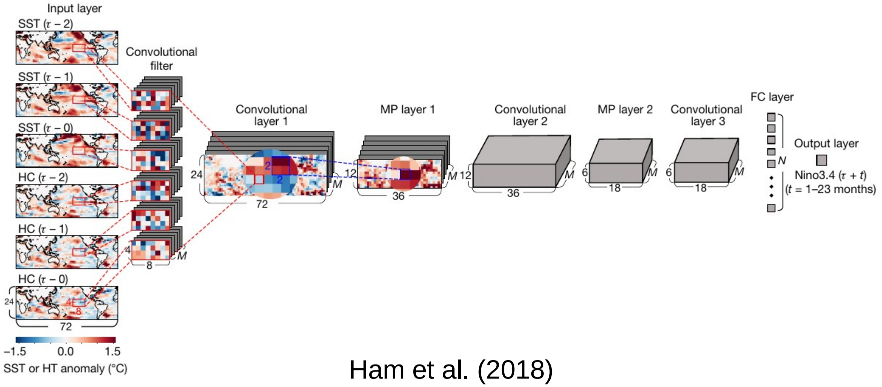
NOAA-PSL Boulder & Universität Tübingen

ENSO Forecast: Statistical (data-driven) models

- **Linear Inverse Model (LIM)**
  - Penland & Matrosova (1994)
  - LIM skill is comparable to NMME (Newman and Sardeshmukh, 2017)

ENSO Forecast: Statistical (data-driven) models

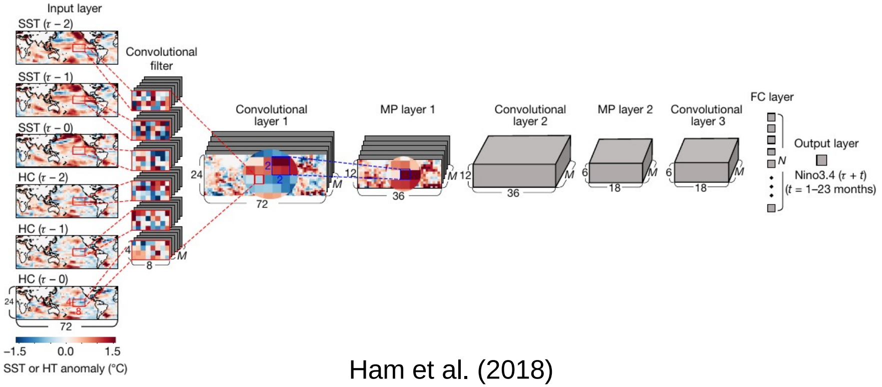
- **Linear Inverse Model (LIM)**
  - Penland & Matrosova (1994)
  - LIM skill is comparable to NMME (Newman and Sardeshmukh, 2017)
- **Neural networks**
  - Convolutional neural networks (Ham et al., 2018 & 2021)
  - ...



Ham et al. (2018)

ENSO Forecast: Statistical (data-driven) models

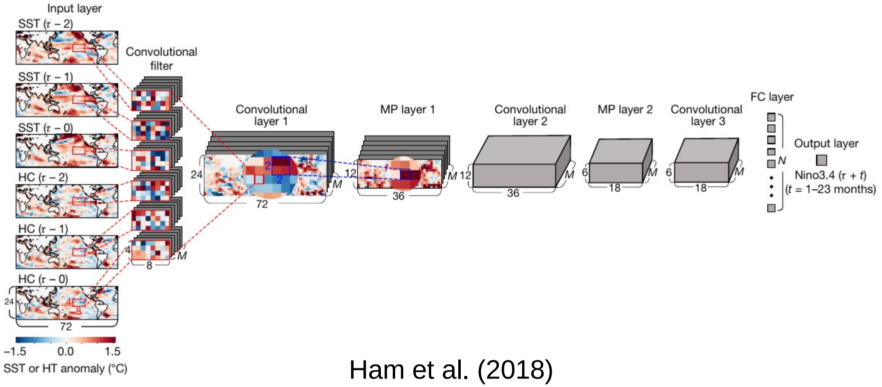
- **Linear Inverse Model (LIM)**
  - Penland & Matrosova (1994)
  - LIM skill is comparable to NMME (Newman and Sardeshmukh, 2017)
- **Neural networks**
  - Convolutional neural networks (Ham et al., 2018 & 2021)
  - ...
- **LIM + Neural networks**
  - Can we improve forecasts?



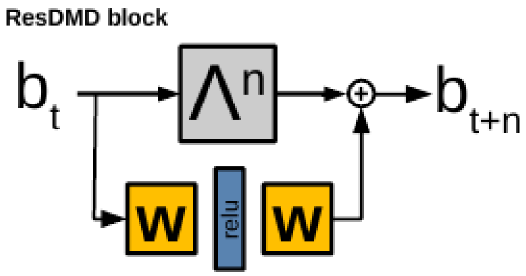
Ham et al. (2018)

ENSO Forecast: Statistical (data-driven) models

- **Linear Inverse Model (LIM)**
  - Penland & Matrosova (1994)
  - LIM skill is comparable to NMME (Newman and Sardeshmukh, 2017)
- **Neural networks**
  - Convolutional neural networks (Ham et al., 2018 & 2021)
  - ...
- **LIM + Neural networks**
  - Can we improve forecasts?



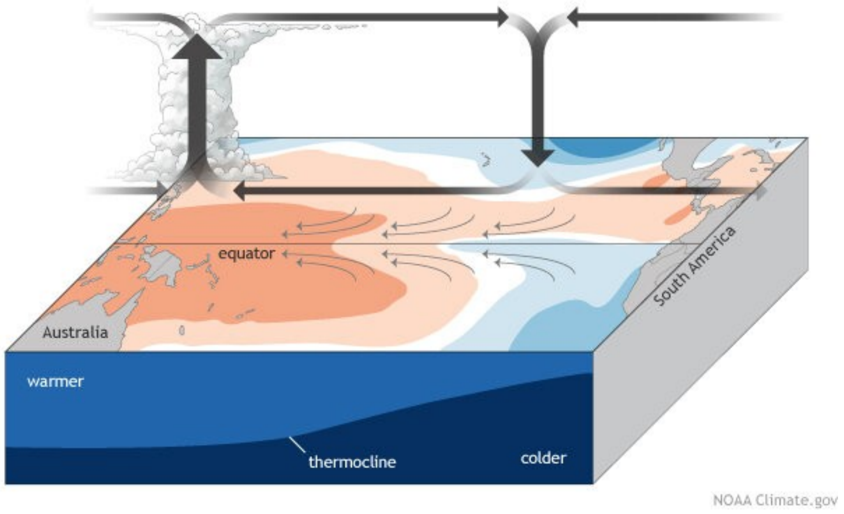
Ham et al. (2018)



Rodrigues et al. (2021)

ENSO Dynamics

$$\frac{dx}{dt} = \underbrace{F\left(x\left(t\right)\right)}_{\text{deterministic}} + \underbrace{G\left(x\left(t\right)\right)\xi}_{\text{stochastic}}$$



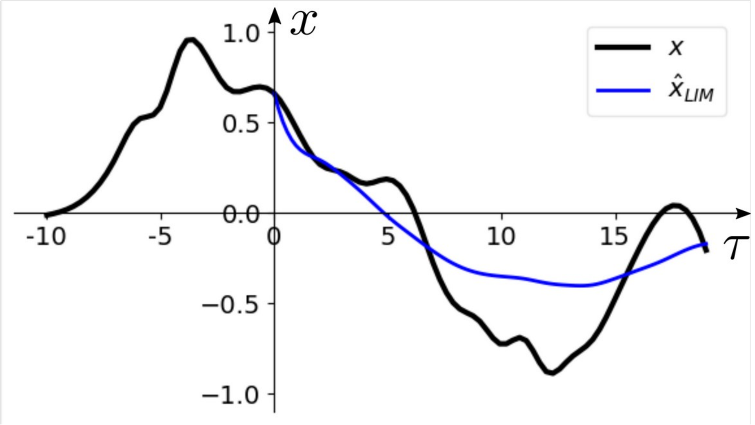
- Slow varying deterministic ocean dynamics
- Ocean is stochastically forced by atmosphere

Penland & Ewald, Phil. Trans. R. Soc. (2008)

Linear Inverse Model (LIM)

$$\frac{dx}{dt} = \underbrace{F\left(x\left(t\right)\right)}_{\text{deterministic}} + \underbrace{G\left(x\left(t\right)\right)\xi}_{\text{stochastic}}$$

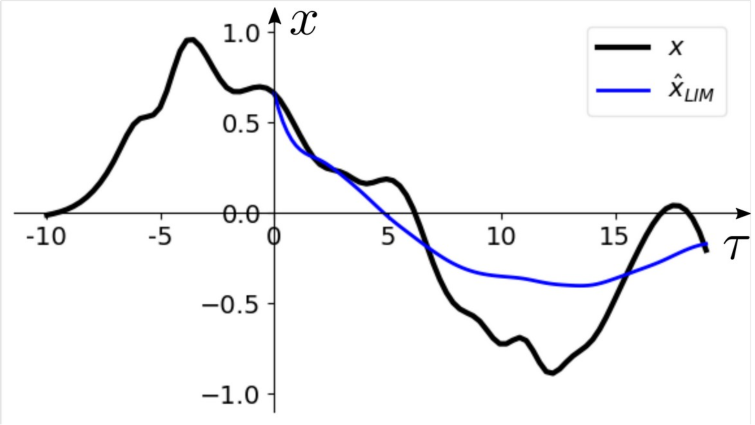
$$\approx L\,x + \xi$$

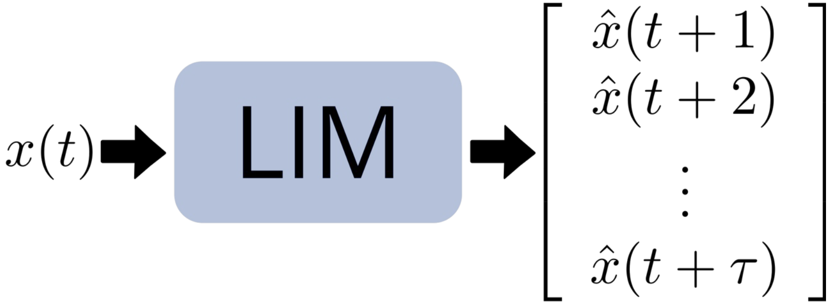


Penland & Matrosova (1994)

Linear Inverse Model (LIM)

$$\frac{dx}{dt} = \underbrace{F\big(x(t)\big)}_{\text{deterministic}} + \underbrace{G\big(x(t)\big)\,\xi}_{\text{stochastic}}$$
$$\approx L\,x + \xi$$



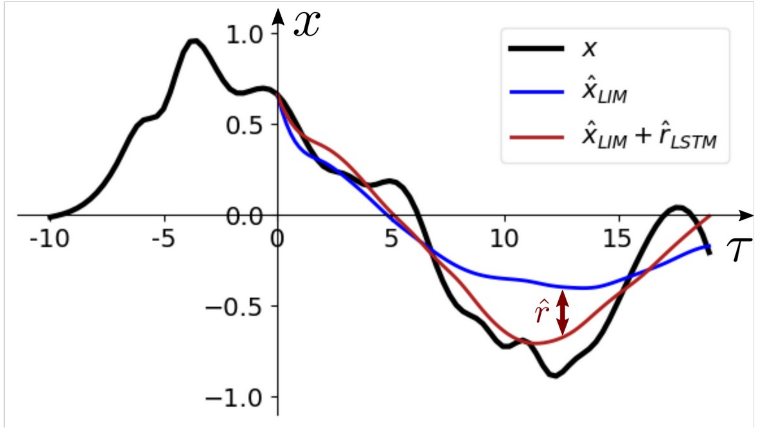


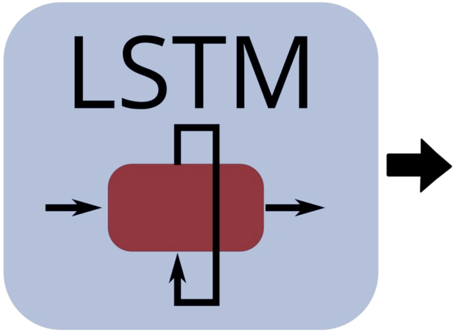
Penland & Matrosova (1994)

Long Short Term Memory (LSTM) Network

$$\frac{dx}{dt} = \underbrace{F\big(x(t)\big)}_{\text{deterministic}} + \underbrace{G\big(x(t)\big)\xi}_{\text{stochastic}}$$

$$\approx L\,x + \underbrace{N(x)}_{} + \xi$$





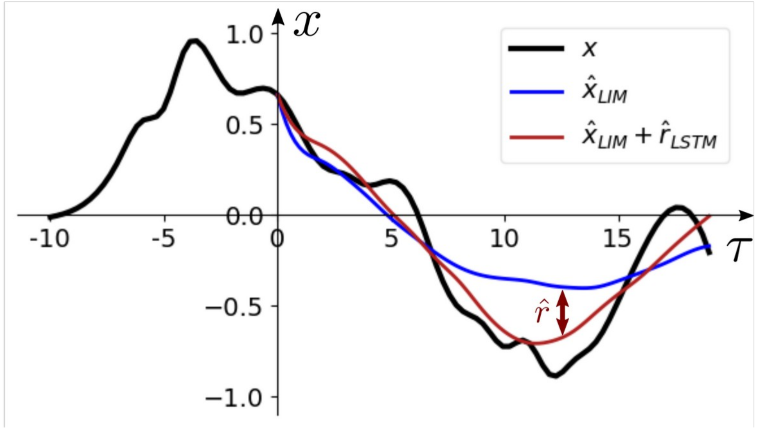
$$\begin{bmatrix} \hat{x}(t+1) + \hat{r}(t+1) \\ \hat{x}(t+2) + \hat{r}(t+2) \\ \vdots \\ \hat{x}(t+\tau) + \hat{r}(t+\tau) \end{bmatrix}$$

Hochreiter & Schmidthuber (1997)

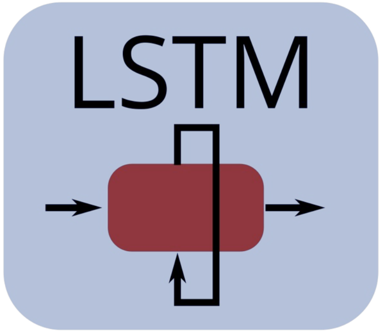
Long Short Term Memory (LSTM) Network

$$\frac{dx}{dt} = \underbrace{F\big(x(t)\big)}_{\text{deterministic}} + \underbrace{G\big(x(t)\big)\,\xi}_{\text{stochastic}}$$

$$\approx Lx + \underbrace{N(x)}_{} + \xi$$



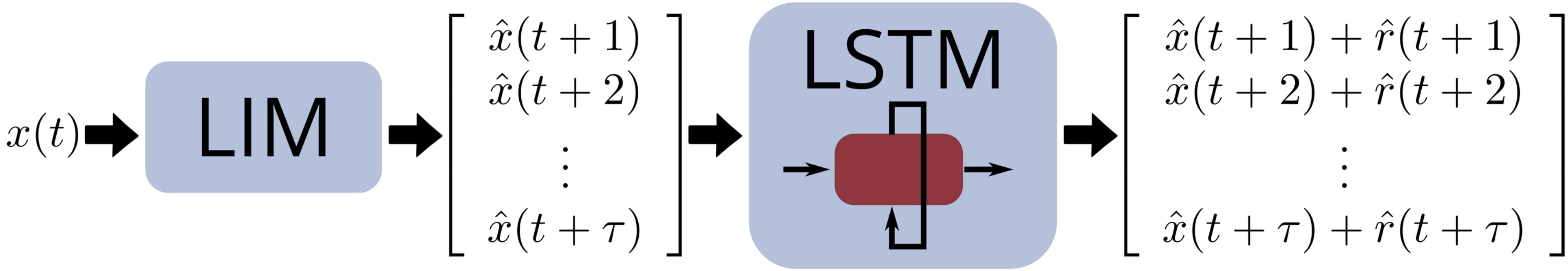
- LSTM is a Recurrent Neural Networks
- Aggregate different time-scales
- Captures nonlinear and non-markovian dynamics



$$\begin{bmatrix} \hat{x}(t+1) + \hat{r}(t+1) \\ \hat{x}(t+2) + \hat{r}(t+2) \\ \vdots \\ \hat{x}(t+\tau) + \hat{r}(t+\tau) \end{bmatrix}$$

Hochreiter & Schmidthuber (1997)

LIM + LSTM



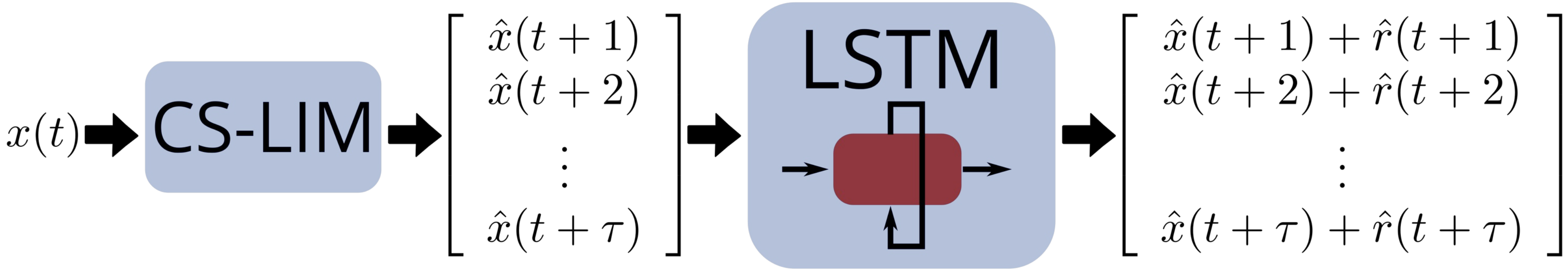
- Training using mean-square error loss:

$$\mathscr{L}\left(x,\hat{x}+\hat{r}\right)=\frac{1}{N}\sum\left(x-\left(\hat{x}+\hat{r}\right)\right)^2$$

Characterizing Nonlinearities in ENSO Dynamics

Jakob Schlör

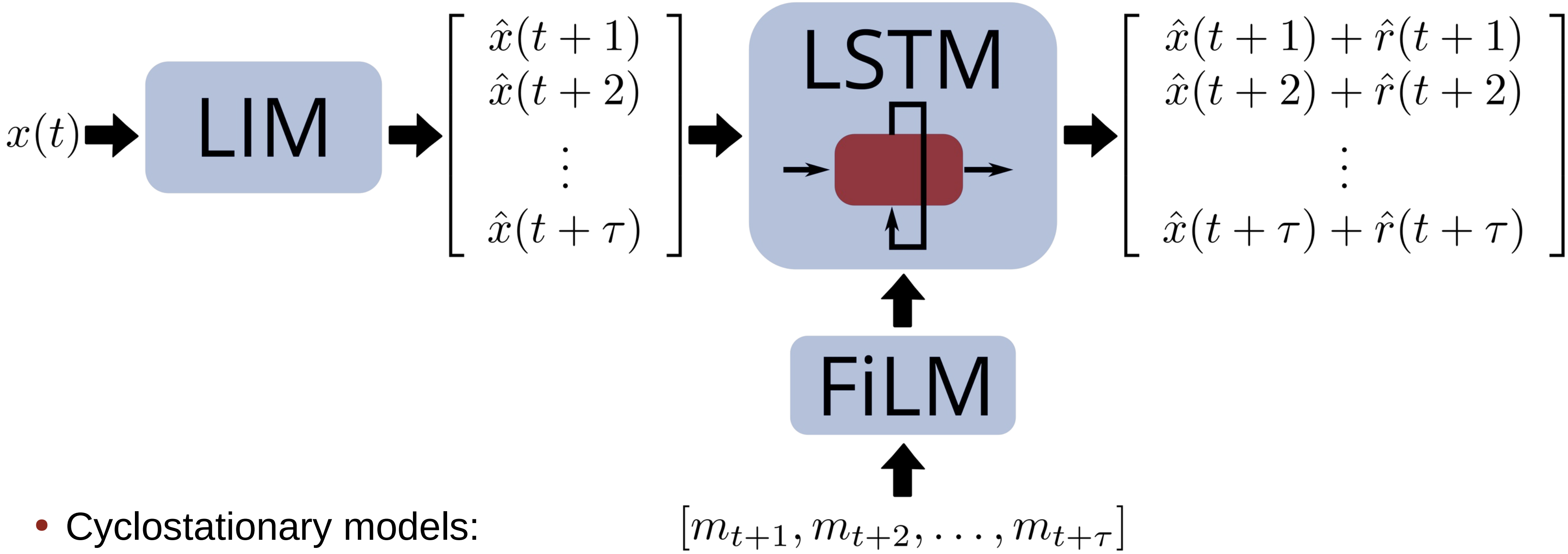
LIM + LSTM



- Cyclostationary models:
  - CS-LIM

Shin et al., JoC (2021)
E. Perez et al., AAAI (2018)

LIM + LSTM

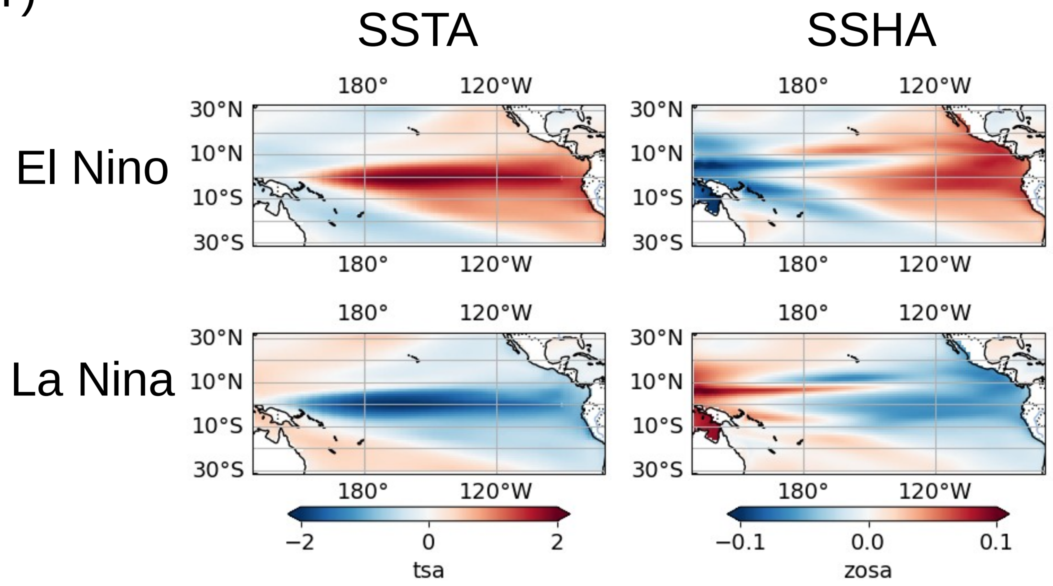


- Cyclostationary models:
  - CS-LIM
  - CS-LSTM: Conditioning on season using Feature-wise Linear Modulation (FiLM)

E. Perez et al., AAAI (2018)

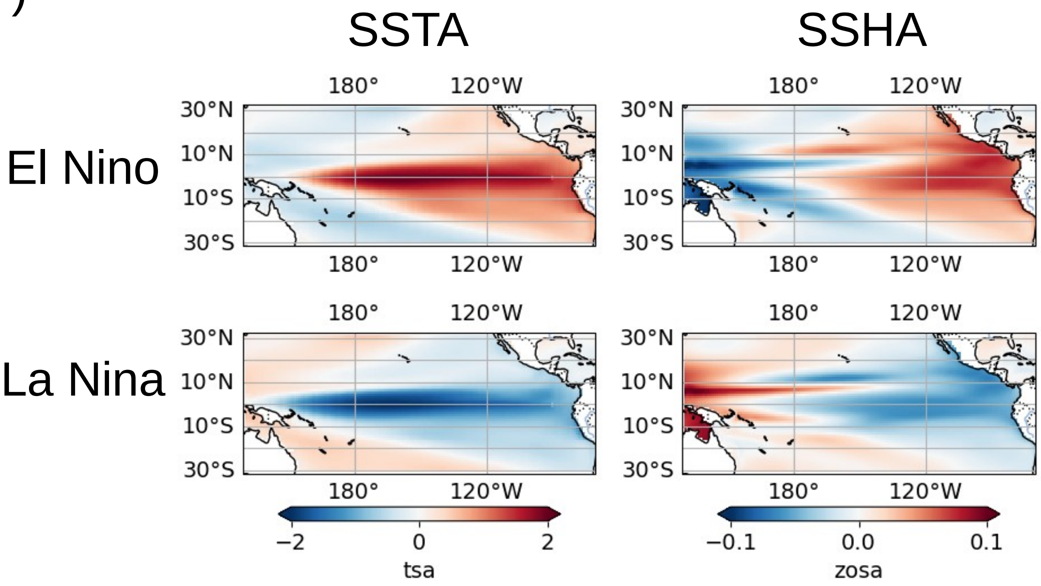
Data: CESM2 Preindustrial Control

- Monthly sea surface temperature (SST) and sea surface height (SSH)
- Tropical Pacific (30°S-30°N, 130°E-70°W)
- Remove monthly mean



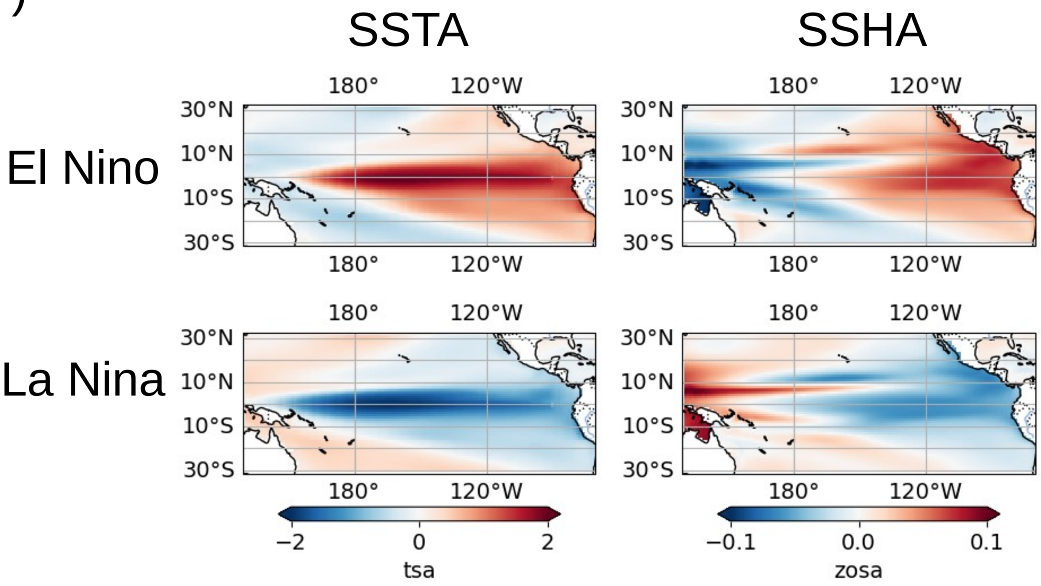
Data: CESM2 Preindustrial Control

- Monthly sea surface temperature (SST) and sea surface height (SSH)
- Tropical Pacific (30°S-30°N, 130°E-70°W)
- Remove monthly mean
- 1200 year of data
- Split data into training (11.5k) and test (2.9k) data



Data: CESM2 Preindustrial Control

- Monthly sea surface temperature (SST) and sea surface height (SSH)
- Tropical Pacific (30°S-30°N, 130°E-70°W)
- Remove monthly mean
- 1200 year of data
- Split data into training (11.5k) and test (2.9k) data
- First 20 PCs of SSTA and 10 PCs of SSHA



Evaluation Metrics

- Evaluation in grid space
- Skill score:

$$\epsilon = 1 - RMSE(\hat{x}, x) / \sigma$$

x : data
$\hat{x}$: prediction
σ : standard deviation of x

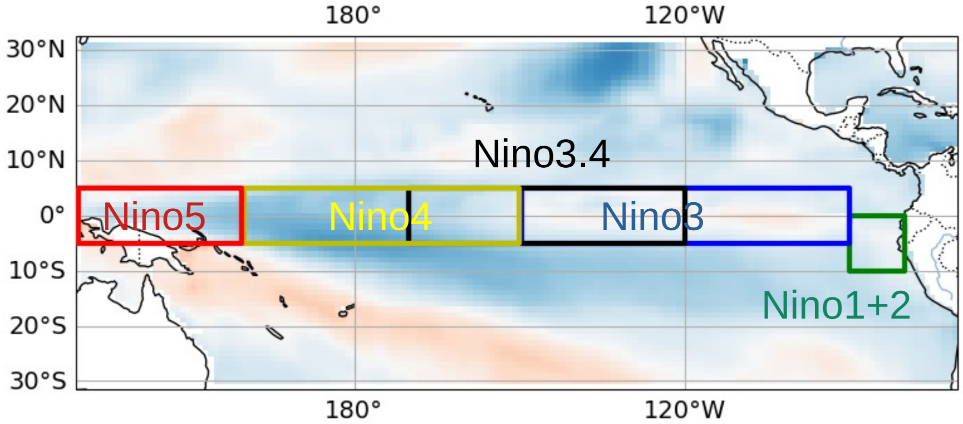
Evaluation Metrics

- Evaluation in grid space
- Skill score:

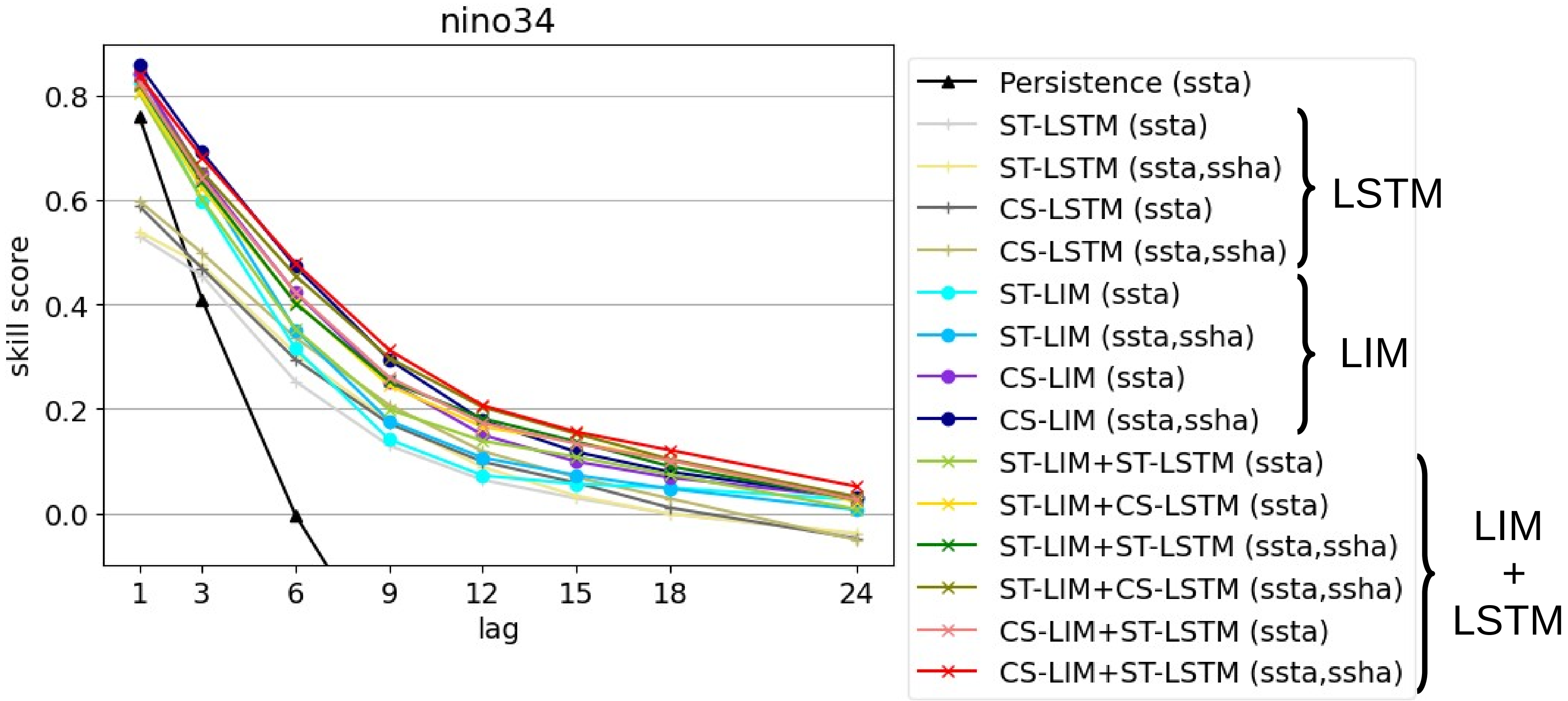
$$\epsilon = 1 - RMSE(\hat{x}, x) / \sigma$$

x : data
$\hat{x}$: prediction
σ : standard deviation of x

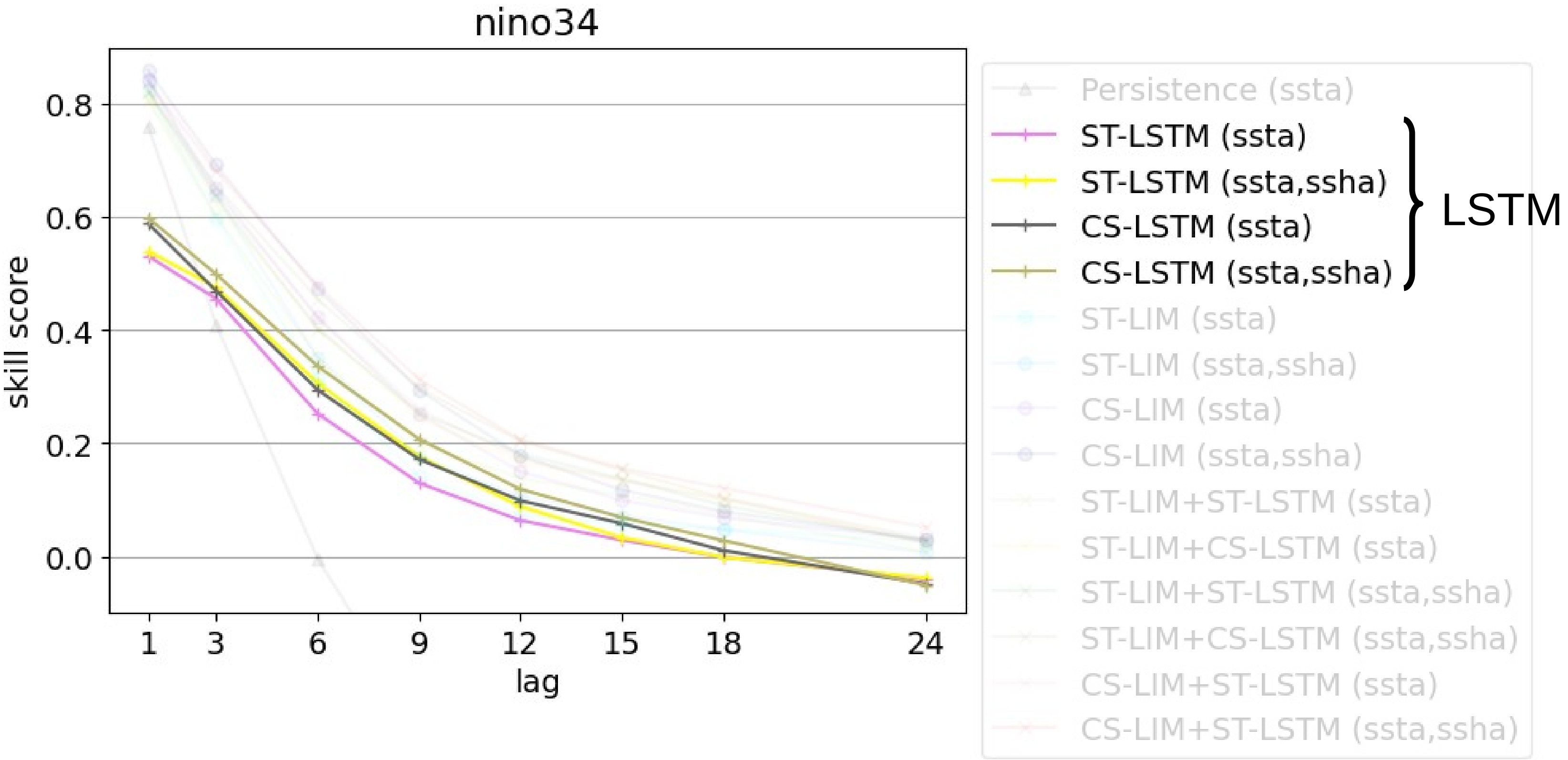
- Averages over Niño-regions



Results: Zoo of experiments



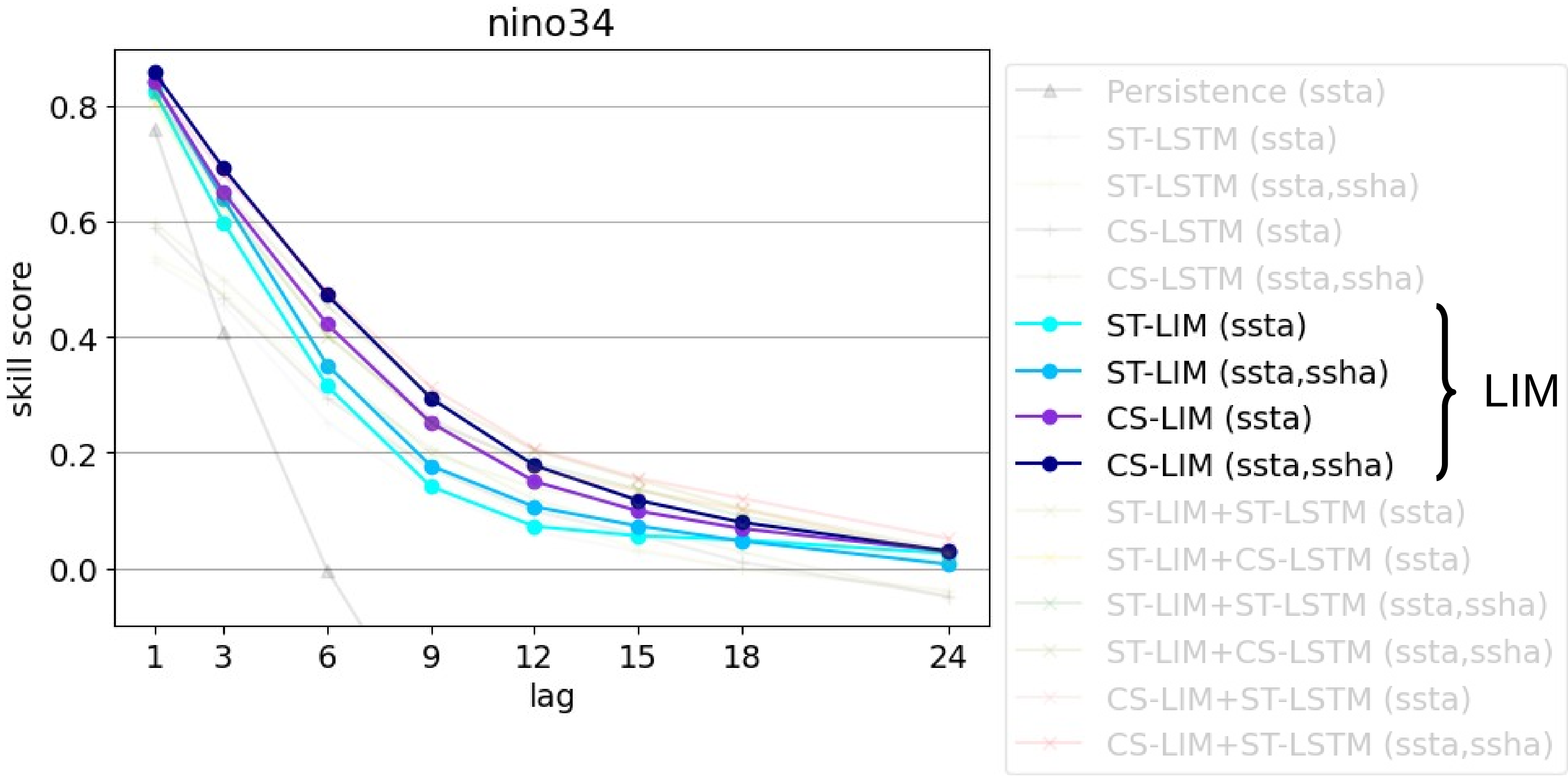
Pure LSTM



Learning the whole dynamics is harder than only residuals

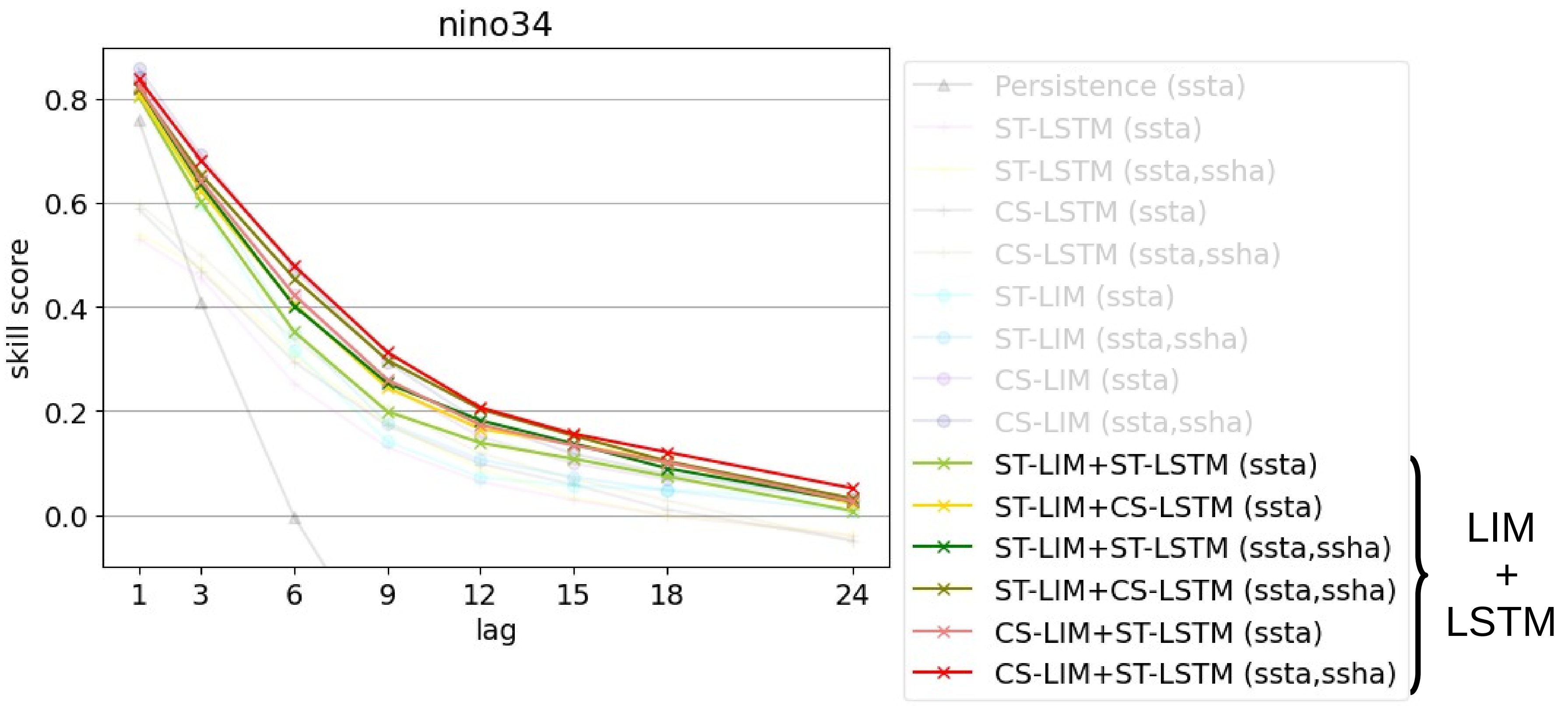
- More data?

LIMs



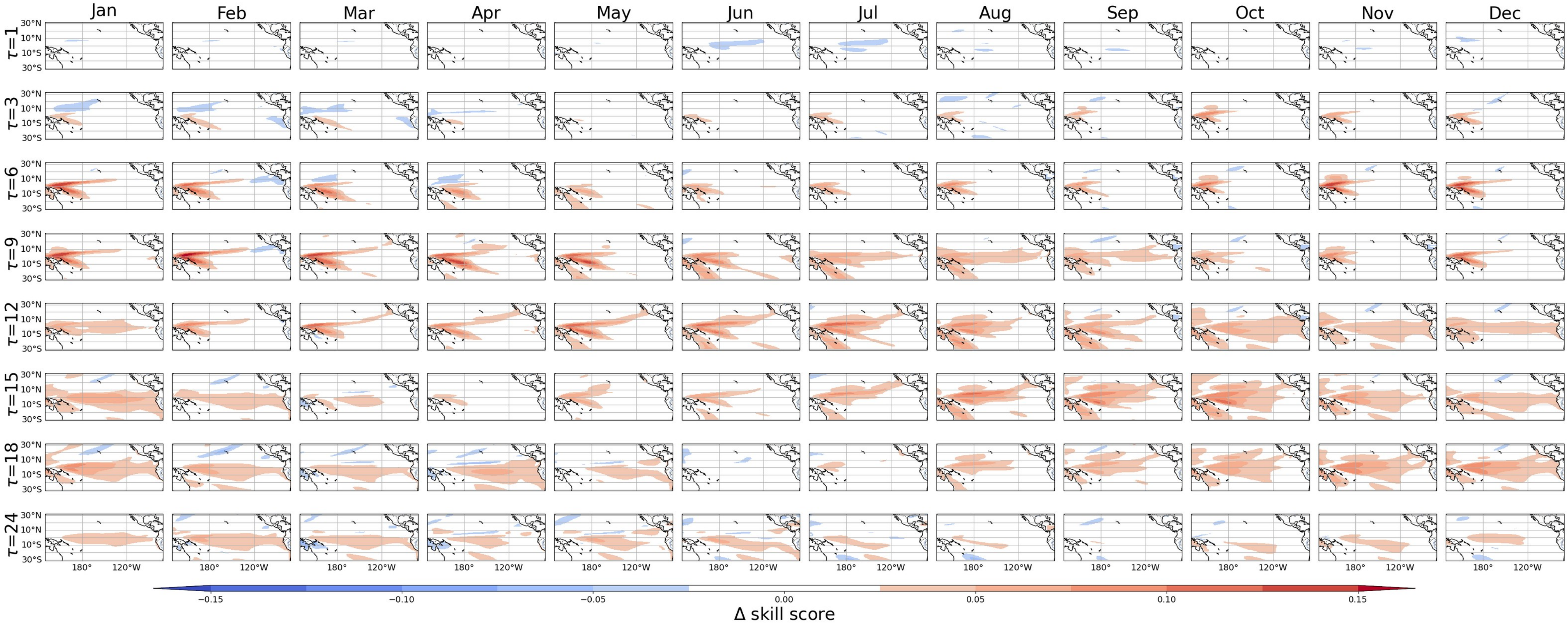
- LIM prediction improves by including seasonality & ocean variable

LIM + LSTM



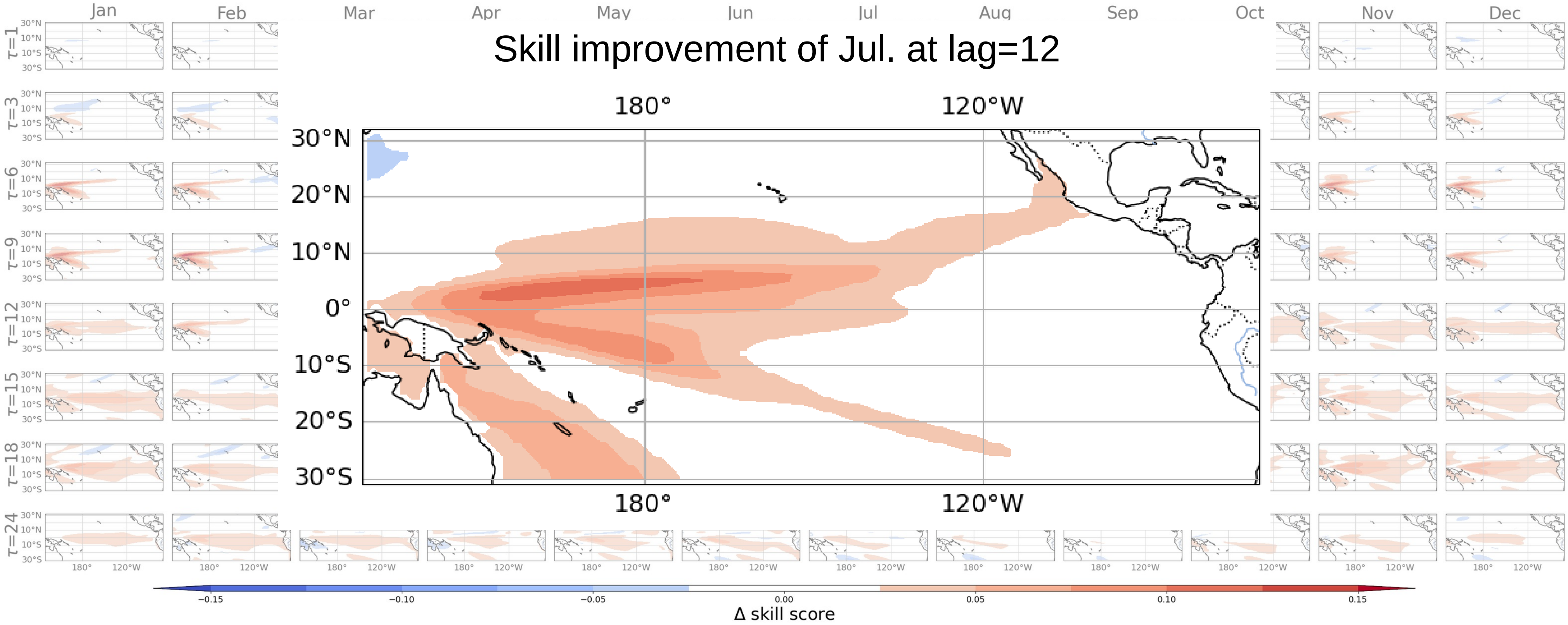
nino34
skill score
lag
0.8
0.6
0.4
0.2
0.0
1
3
6
9
12
15
18
24
Persistence (ssta)
ST-LSTM (ssta)
ST-LSTM (ssta,ssha)
CS-LSTM (ssta)
CS-LSTM (ssta,ssha)
ST-LIM (ssta)
ST-LIM (ssta,ssha)
CS-LIM (ssta)
CS-LIM (ssta,ssha)
ST-LIM+ST-LSTM (ssta)
ST-LIM+CS-LSTM (ssta)
ST-LIM+ST-LSTM (ssta,ssha)
ST-LIM+CS-LSTM (ssta,ssha)
CS-LIM+ST-LSTM (ssta)
CS-LIM+ST-LSTM (ssta,ssha)
LIM
+
LSTM

What dynamics does the LSTM pick up?



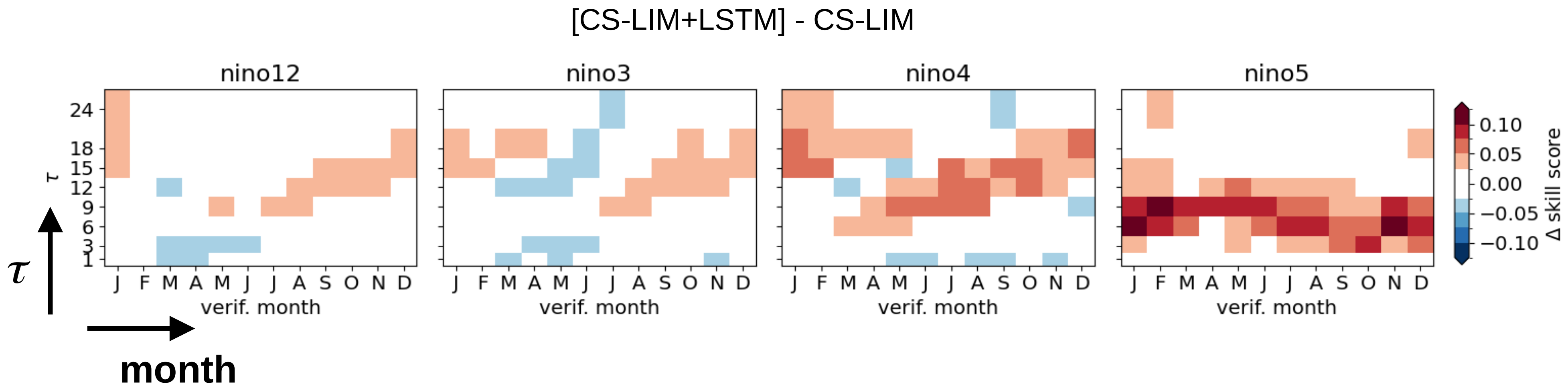
Skill improvement by LSTM, i.e. [CS-LIM+LSTM] - CS-LIM

What dynamics does the LSTM pick up?

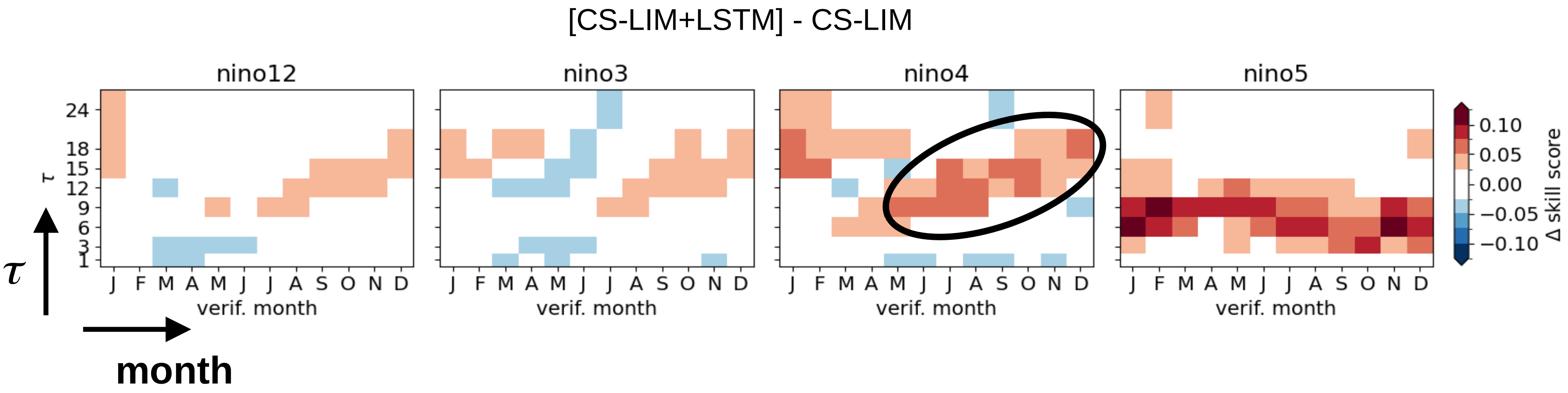


Skill improvement by LSTM, i.e. [CS-LIM+LSTM] - CS-LIM

What dynamics does the LSTM pick up?

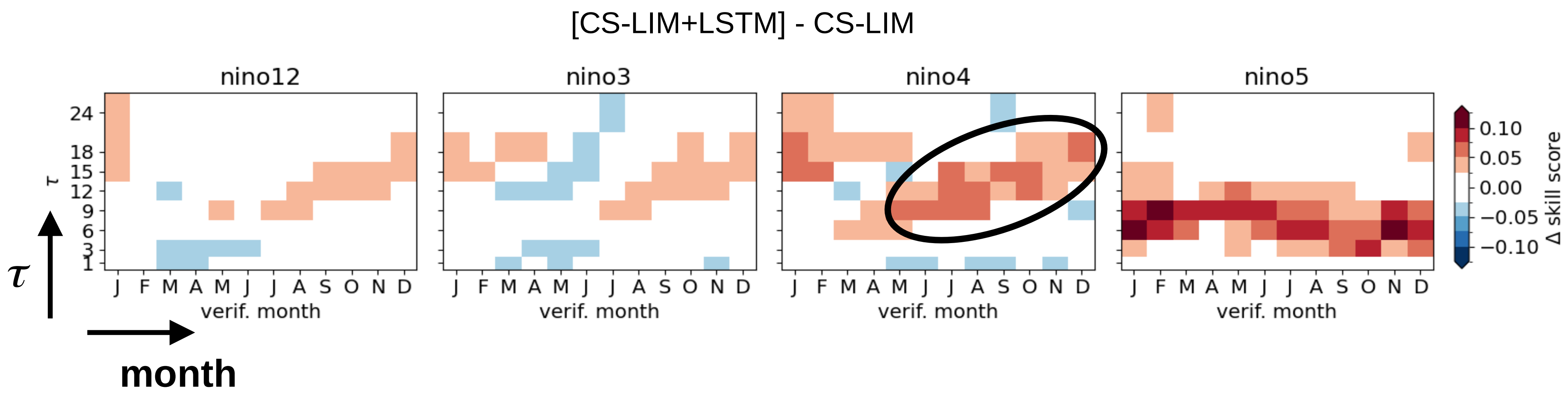


What dynamics does the LSTM pick up?



- LSTM improves forecast from 9-18 months in the Western Pacific
- Improved forecasts are initialized in July – December

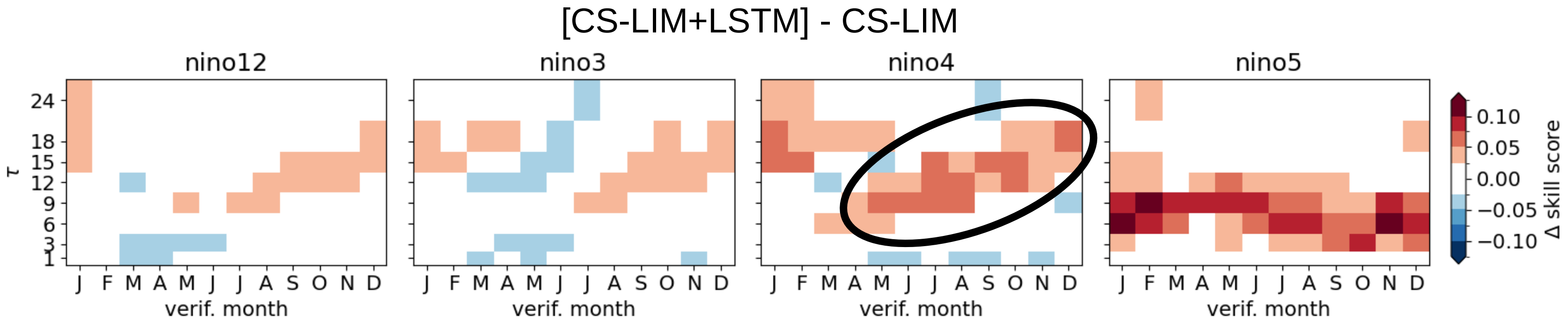
What dynamics does the LSTM pick up?



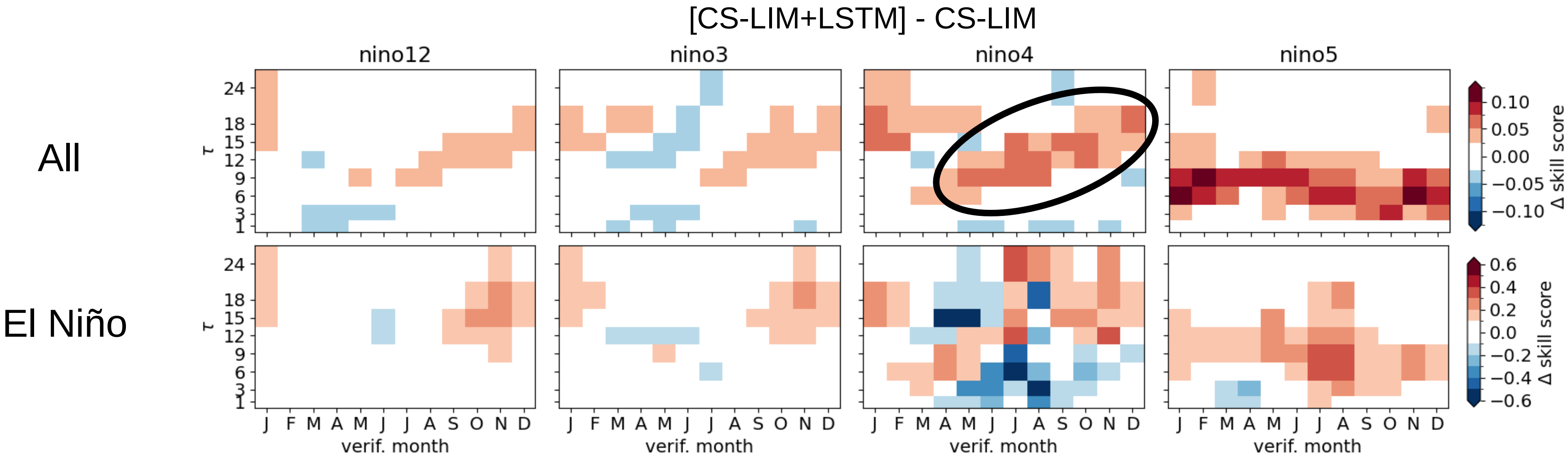
- LSTM improves forecast from 9-18 months in the Western Pacific
- Improved forecasts are initialized in July – December
- We know it is NOT:
  - Seasonal cycle (linear)
  - Ocean memory

Hypothesis: El Niño - La Niña asymmetry

All



Hypothesis: El Niño - La Niña asymmetry

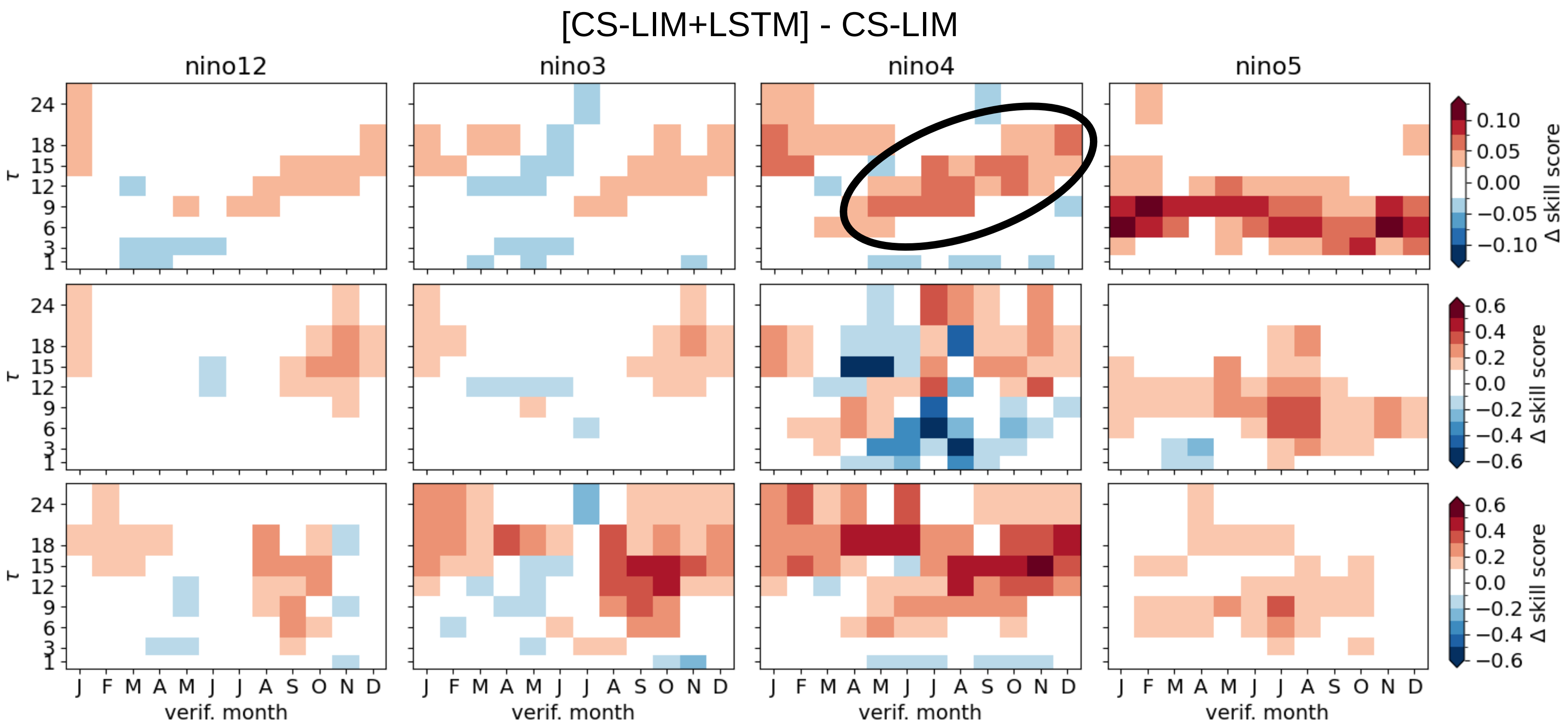


Hypothesis: El Niño - La Niña asymmetry

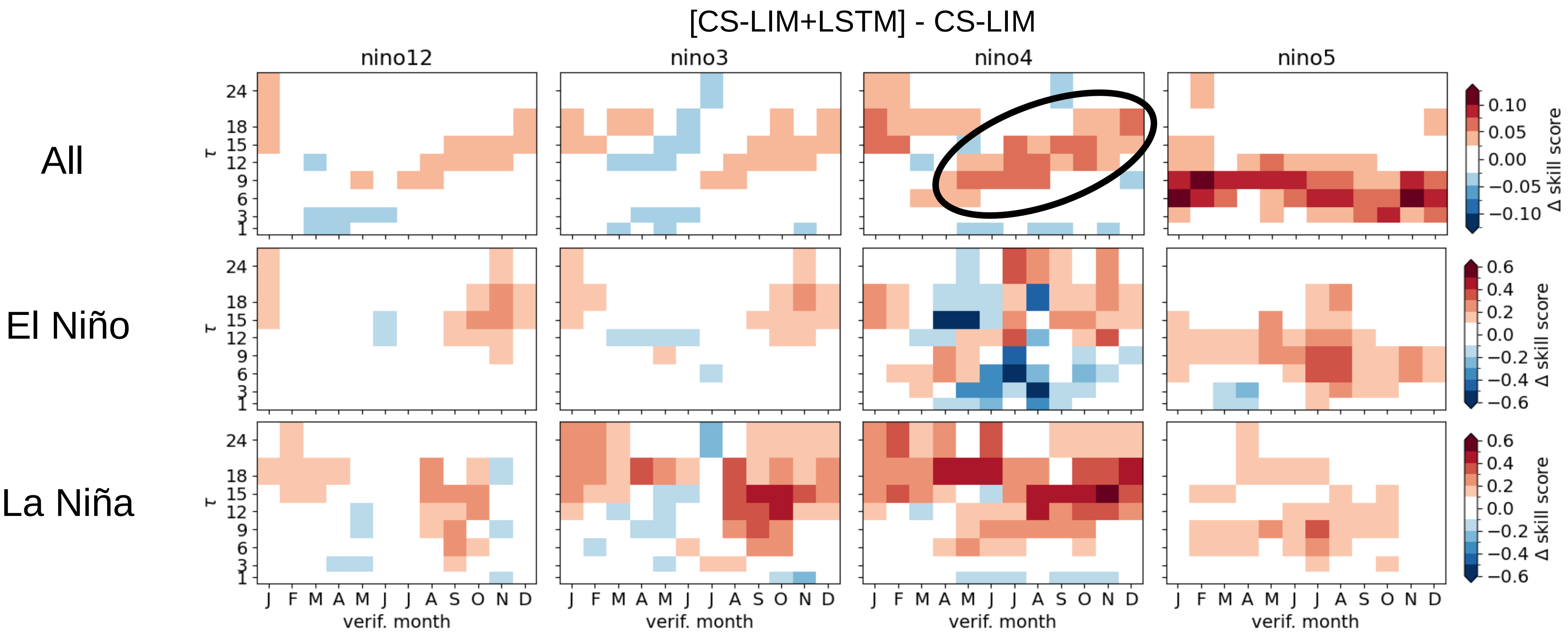
All

El Niño

La Niña



Hypothesis: El Niño - La Niña asymmetry



Skill improvement by LSTM can be partially explained by ENSO asymmetry.

Conclusion

1) Limited amount of data makes S2S prediction hard for LSTM

Conclusion

1) Limited amount of data makes S2S prediction hard for LSTM
2) Non-linearities/ non-markovianity improves prediction for > 9 month

Conclusion

1) Limited amount of data makes S2S prediction hard for LSTM
2) Non-linearities/ non-markovianity improves prediction for > 9 month
3) Predictable nonlinearities are likely due to EN/LN asymmetry

Conclusion

1) Limited amount of data makes S2S prediction hard for LSTM
2) Non-linearities/ non-markovianity improves prediction for > 9 month
3) Predictable nonlinearities are likely due to EN/LN asymmetry

Outlook

- Disentangle nonlinearities from markovianity by idealized experiments
- Apply on observational data

Thank you!

Antonietta Capotondi

Matthew Newmann

Bedartha Goswami

machineclimate.de
@schloer_jakob
jakob.schloer@uni-tuebingen.de

Conclusion

1) Limited amount of data makes S2S prediction hard for LSTM
2) Non-linearities/ non-markovianity improves prediction for > 9 month
3) Predictable nonlinearities are likely due to EN/LN asymmetry

Outlook

- Disentangle nonlinearities from markovianity by idealized experiments
- Apply on observational data

machineclimate.de

@schloer_jakob

jakob.schloer@uni-tuebingen.de

Appendix

Linear Inverse Models

$$\boxed{\frac{dx}{dt} = Lx + \xi}$$

τ : lag time
ξ : white noise forcing
$C(0)$: data covariance
$C(\tau)$: lagged covariance

Stationary LIM:

$$\hat{x}(t+\tau) \sim N\left(\mathbf{exp}\left(\boldsymbol{L}^{ST}\boldsymbol{\tau}\right)\boldsymbol{x}(\boldsymbol{t}), \Sigma^{ST}\right)$$

with same linear operator for all times: $L^{ST} = \tau^{-1}\ln\left[C(\tau)C(0)^{-1}\right]$

Cyclostationary LIM:

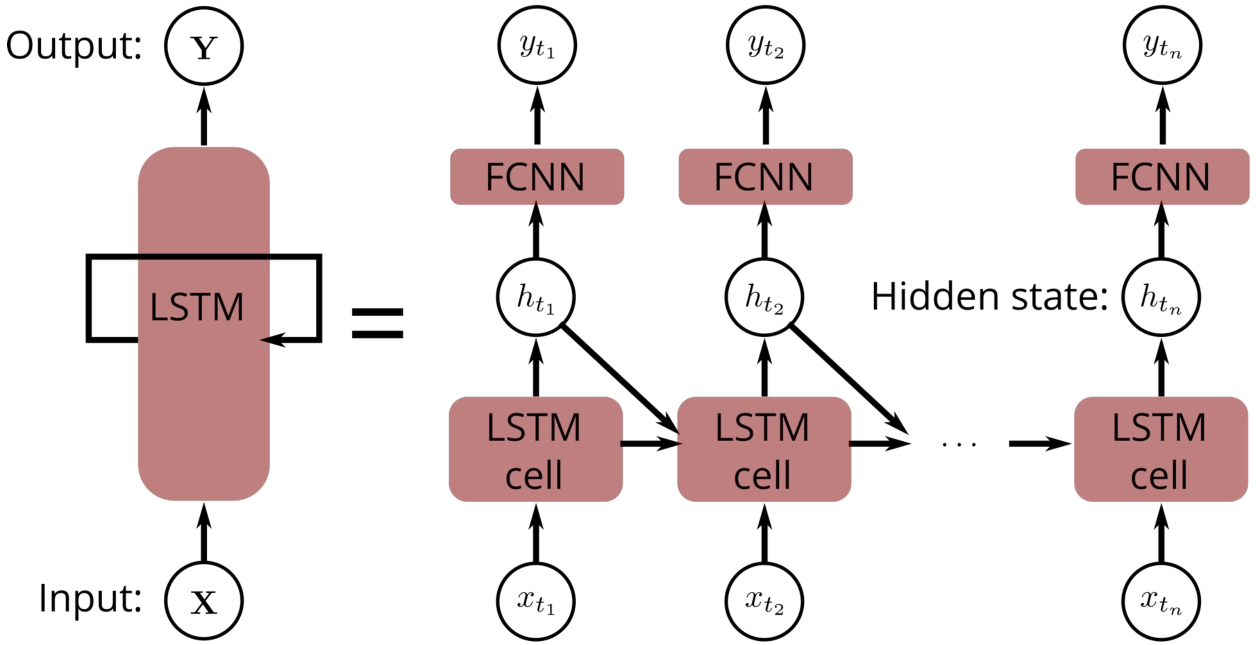
$$\hat{x}_j(t+\tau) \sim N\left(\mathbf{exp}\left(\Sigma_{n=0}^{\tau-1}\boldsymbol{L}_{j-n}^{CS}\right)\boldsymbol{x}_{j-\tau}(\boldsymbol{t}), \Sigma_{j-\tau\rightarrow j}^{CS}\right)$$

with linear operator for each month $j = 1, \ldots, 12$: $L_j^{CS} = \tau_0^{-1}\ln\left[C_j(1)C_j(0)^{-1}\right]$

example Jan $\rightarrow$ Mar: $\hat{x}_3(t+2) \sim N\left(\mathbf{exp}\left(\boldsymbol{L}_{2\rightarrow3}^{CS} + \boldsymbol{L}_{1\rightarrow2}^{CS}\right)\boldsymbol{x}_1(\boldsymbol{t}), \Sigma_{1\rightarrow3}^{CS}\right)$

Shin et al., JoC (2021)

Long Short Term Memory (LSTM) Network



- LSTM is a special kind of Recurrent Neural Networks
- Basic idea: Different time-scales (memory) should be captured by different variables
- Captures nonlinear and non-markovian dynamics

Hochreiter & Schmidthuber (1997)

Long Short Term Memory (LSTM) Network

Cell state: $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

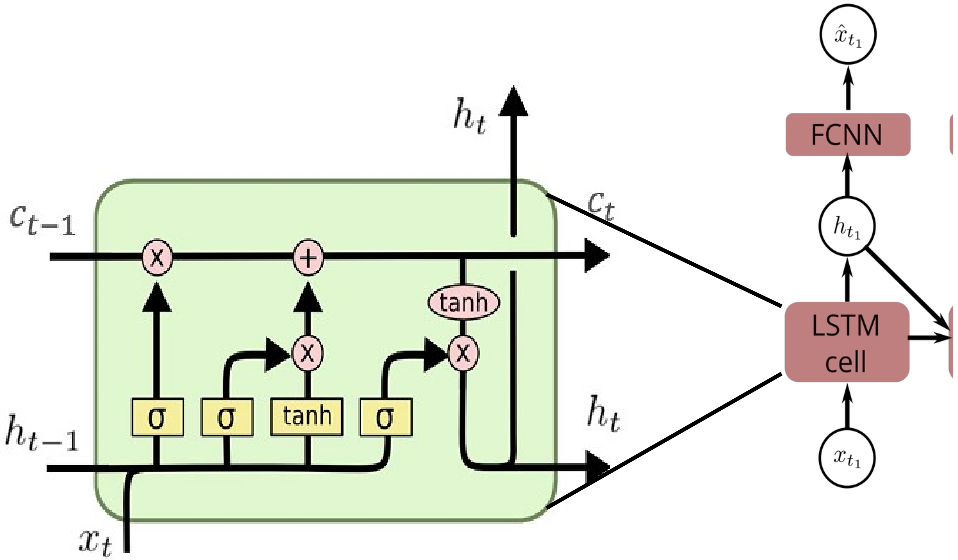
Forget gate: $f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$

Input gate: $i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$

$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$

Hidden state: $o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$

$h_t = o_t * \tanh\left(C_t\right)$



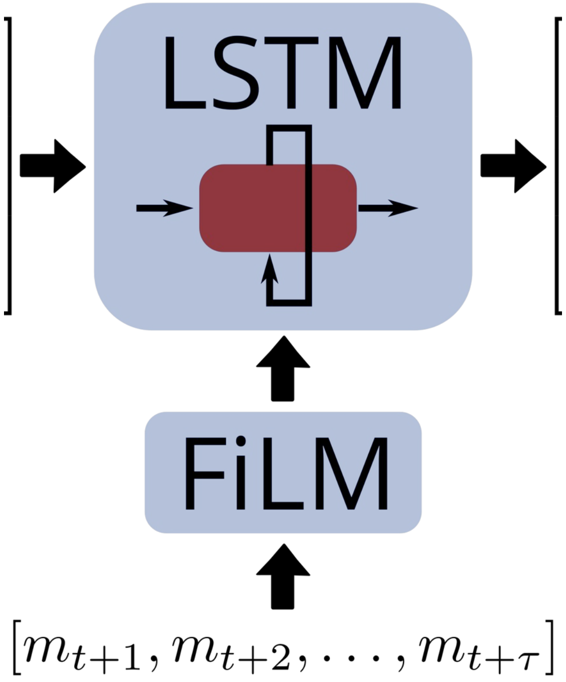
Hochreiter & Schmidthuber (1997)
https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Feature-wise Linear Modulation (FiLM)

- FiLM is used for conditioning NN on external input
- Feature-wise affine transformation of input,
  i.e. scaling and shifting applied element-wise

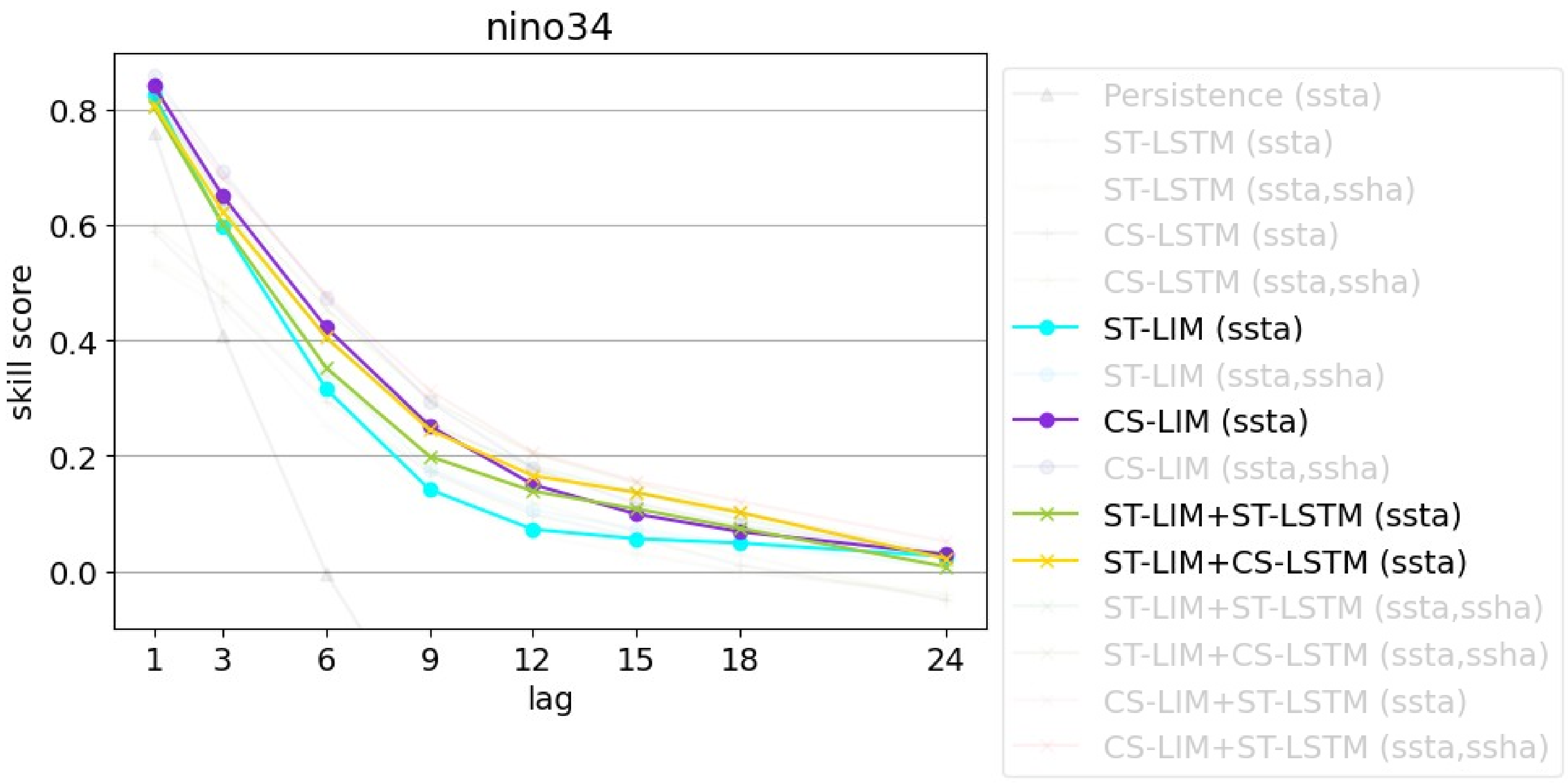
$$\mathrm{FiLM}(x) = \gamma(z) * x + \beta(z)$$

$$\begin{aligned} z &: \text{conditional input} \\ \gamma(z) &: \text{scaling network} \\ \beta(z) &: \text{bias network} \end{aligned}$$

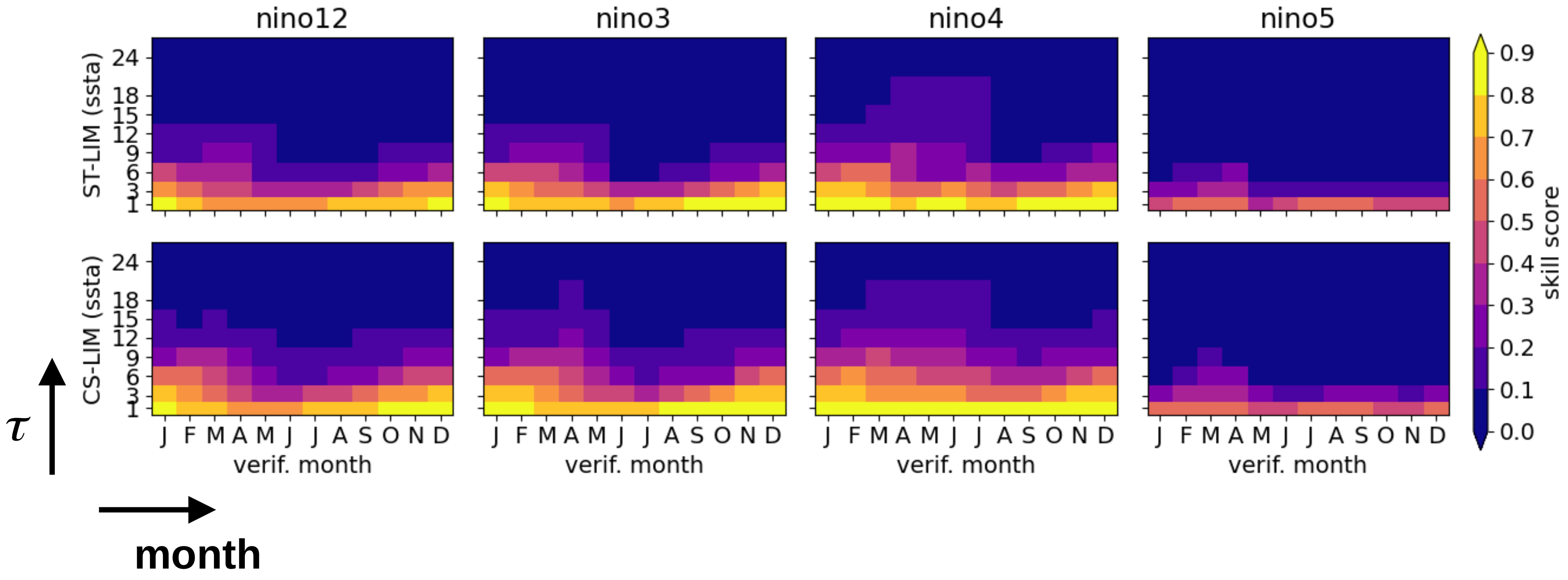


https://distill.pub/2018/feature-wise-transformations/
Perez et al., FiLM(2017)

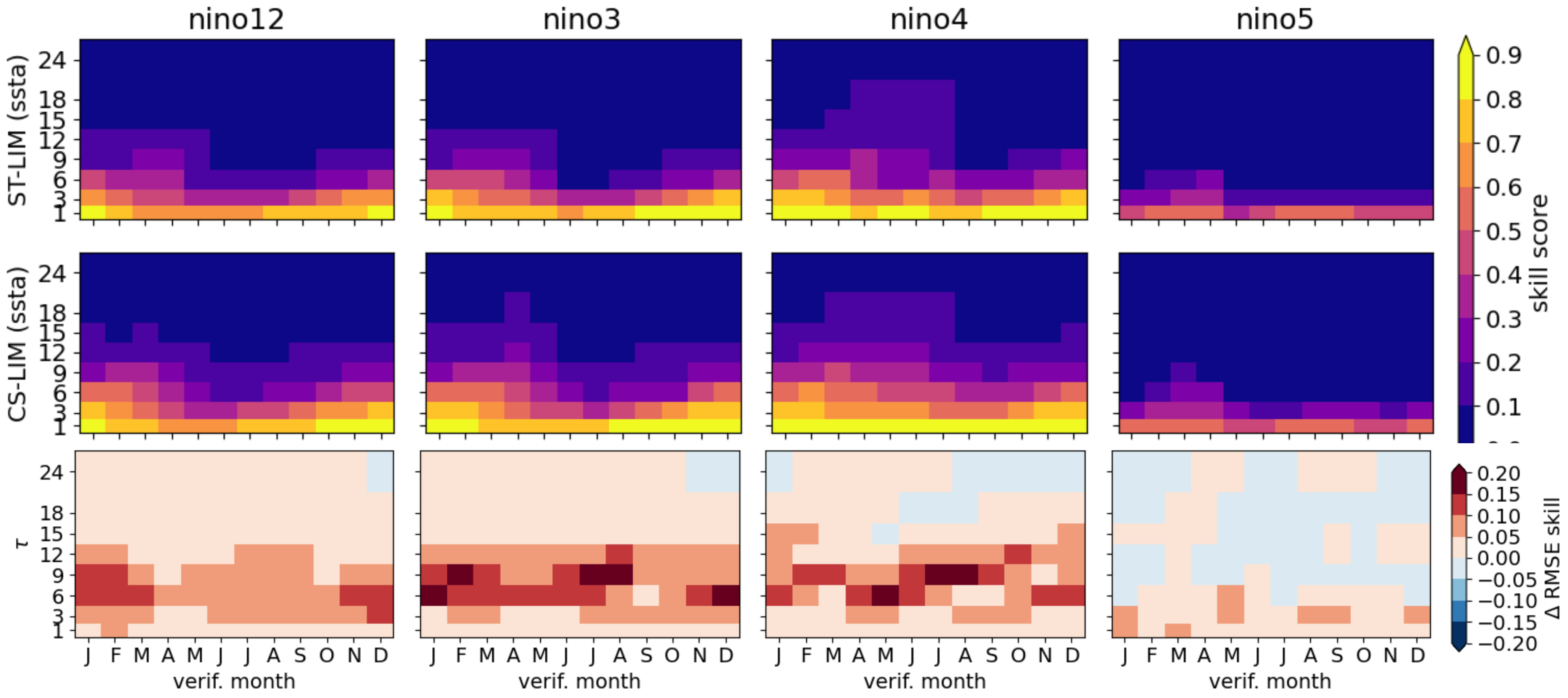
The Seasonal Cycle



The Seasonal Cycle



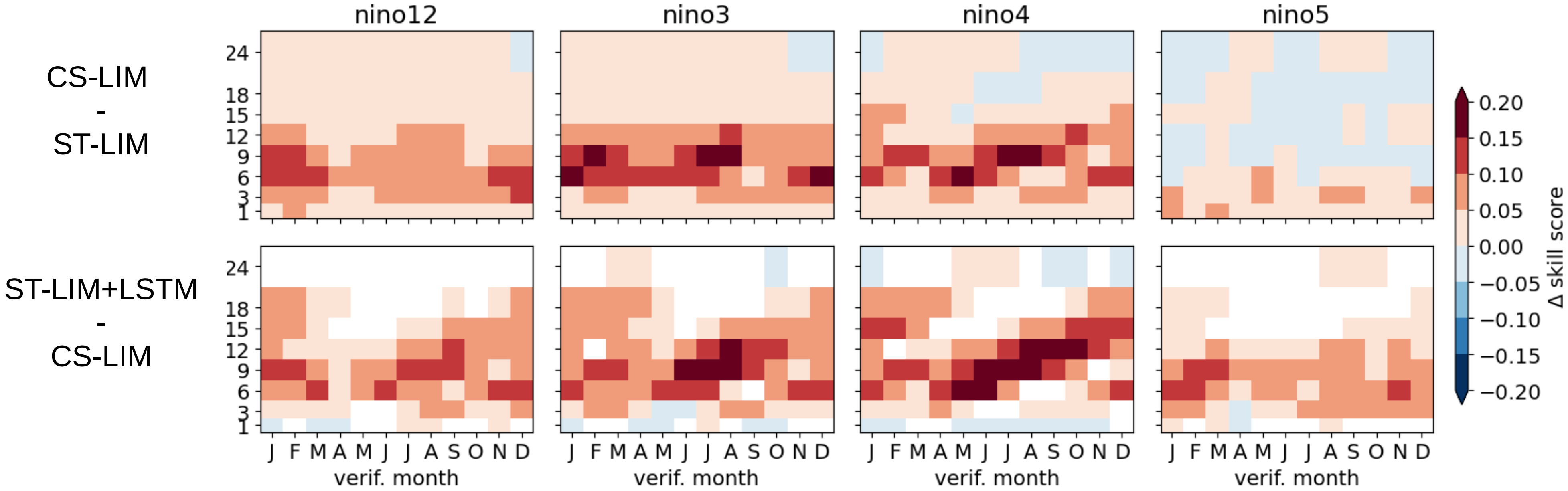
The Seasonal Cycle



CS-LIM
-
ST-LIM

- Including seasonality (linearly) improves forecast between 3-9 month

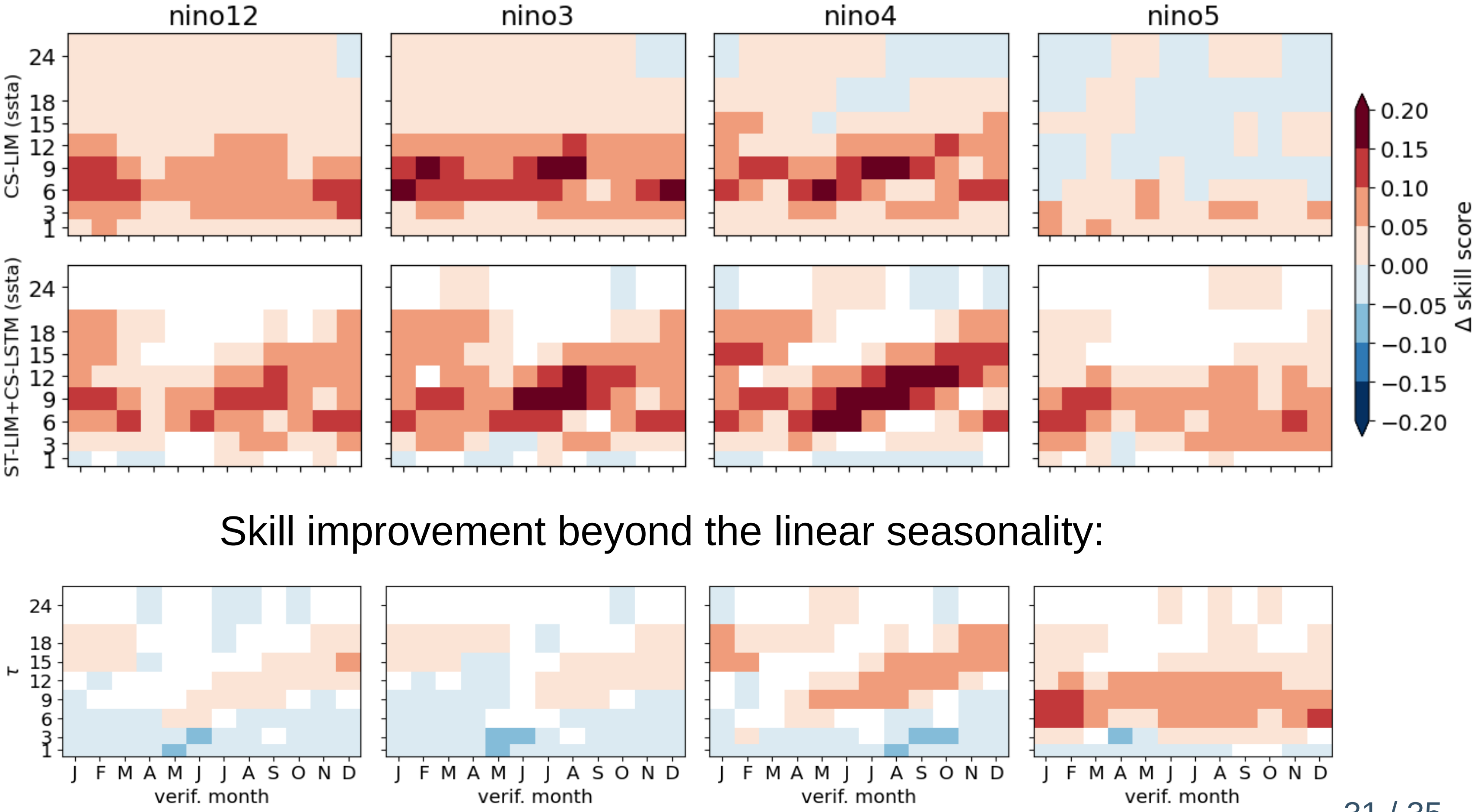
The Seasonal Cycle



The Seasonal Cycle: Skill improvement wrt. ST-LIM

Linear seasonality

?



Skill improvement beyond the linear seasonality:

(ST-LIM+LSTM)
-
CS-LIM

The Ocean Memory

$$\frac{dx}{dt} = \underbrace{F(x(t))}_{\text{deterministic}} + \underbrace{G(x(t))\zeta(t)}_{\text{stochastic}}$$

$$\approx Lx + \underbrace{\mathcal{N}(x)} + \zeta$$

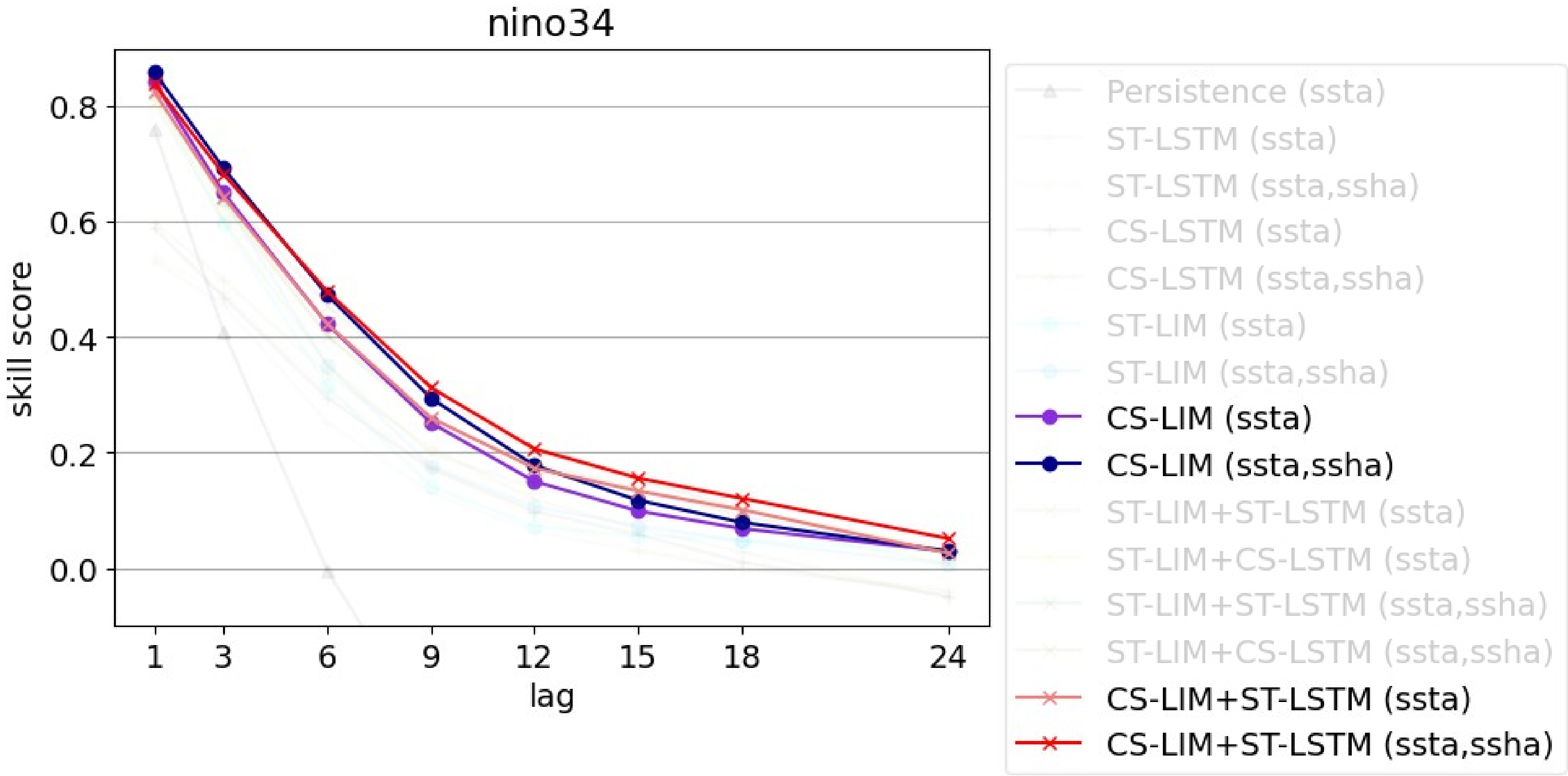
➡ Seasonality: $L_c x + \mathcal{N}_c(x)$

➡ Memory effect:

- Non-Markovianity: $\mathcal{N}_m\left(x(t), x(t-1), \ldots\right)$
- Ocean variables: $x = (\text{ssta}, \text{ssha})$

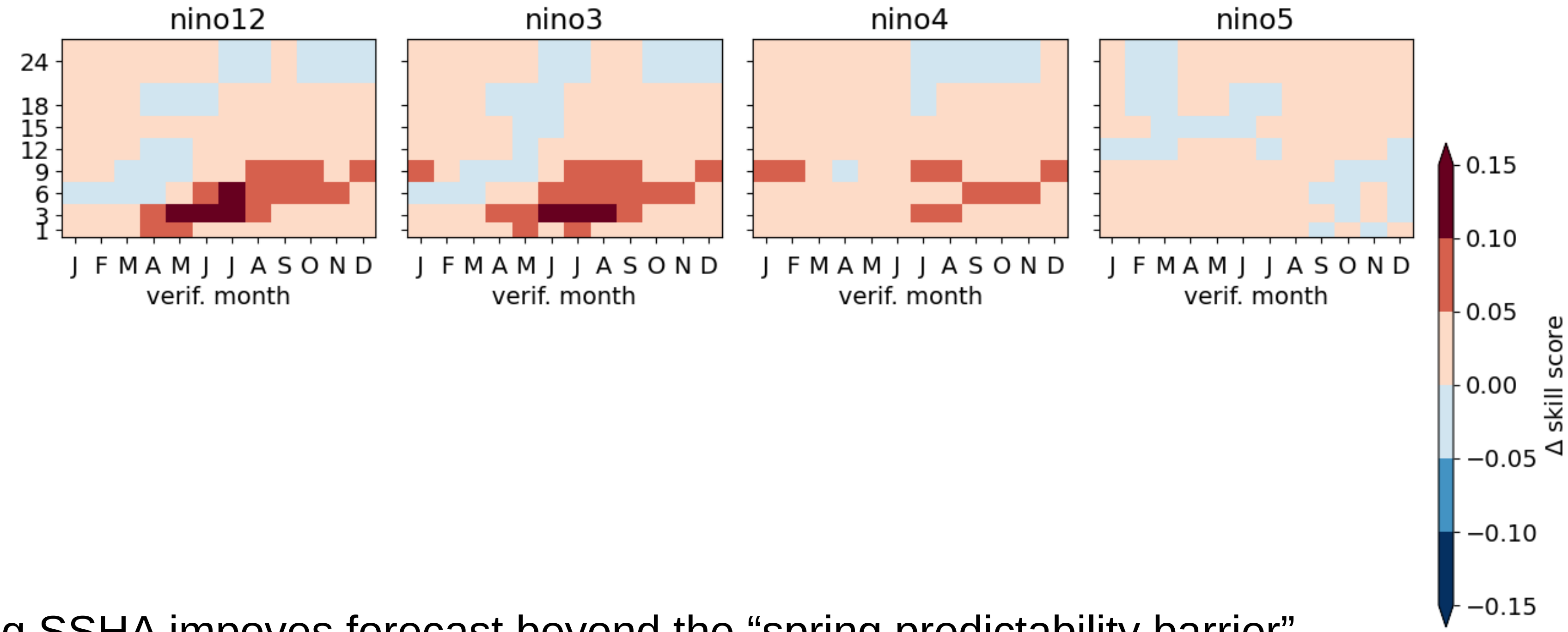
➡ Other nonlinearities

The Ocean Memory



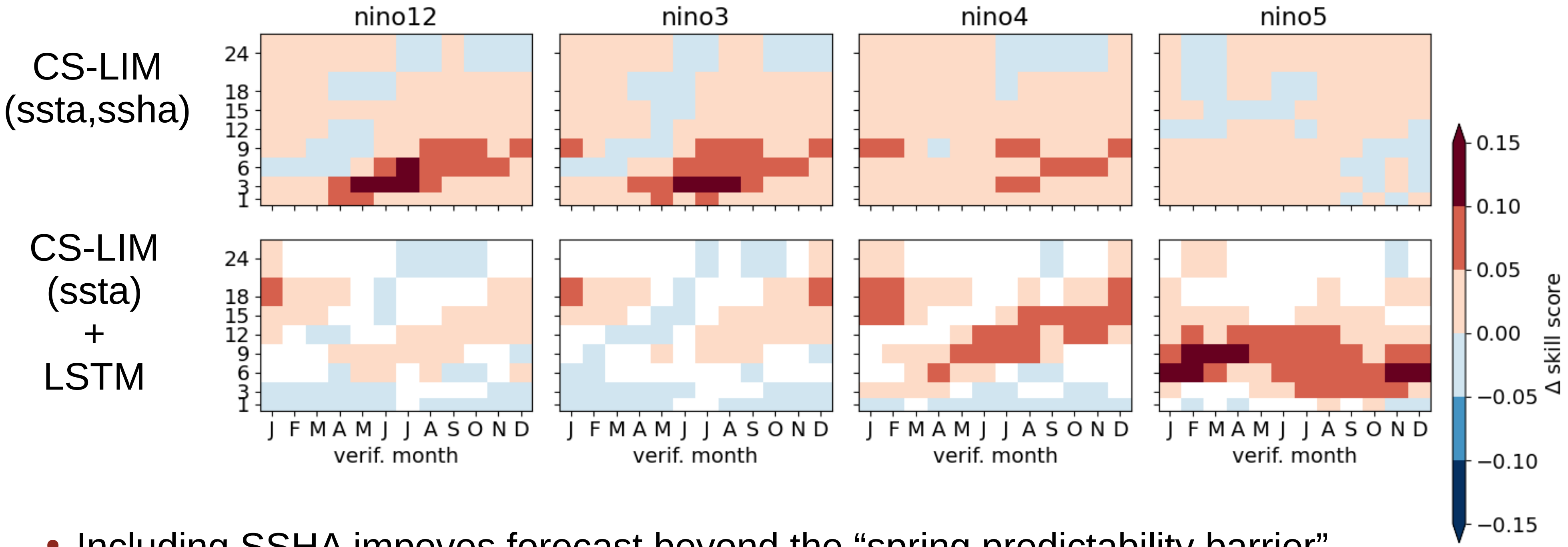
The Ocean Memory: Skill improvement beyond CS-LIM (ssta)

CS-LIM (ssta,ssha)



- Including SSHA impoves forecast beyond the “spring predictability barrier” from 3-9 months mainly in the Eastern Pacific

The Ocean Memory: Skill improvement beyond CS-LIM (ssta)



- Including SSHA impoves forecast beyond the “spring predictability barrier” from 3-9 months mainly in the Eastern Pacific
- LSTM improves forecast from 9-18 months in the Western Pacific