

# Dynamic Time Warping como metodologia para treinamento de modelos de previsão para Temperatura da Superfície Oceânica

Maiser José Alves Oliva, 6309717  
Matheus Victal Cerqueira, 10276661  
Vitor Amorim Fróis, 12543440

Aprendizado de Máquina para Séries Temporais  
Professor: Diego Furtado Silva

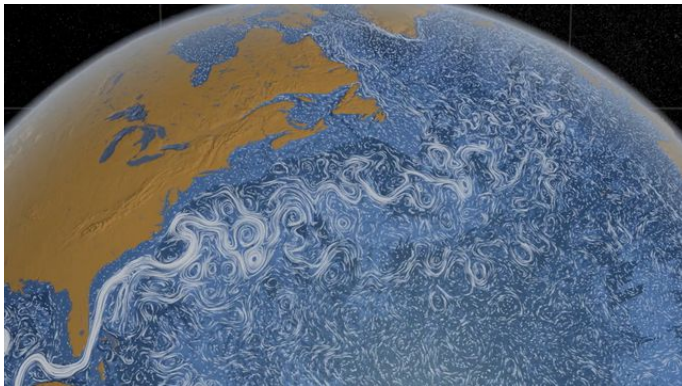
2023

# Sumário

- ▶ Temperatura da Superfície Oceânica
- ▶ SARIMA
- ▶ LSTM
- ▶ DS (DTW + SVR)
- ▶ Comparações de performance RMSE / MAPE
- ▶ SARIMAX walk-forward

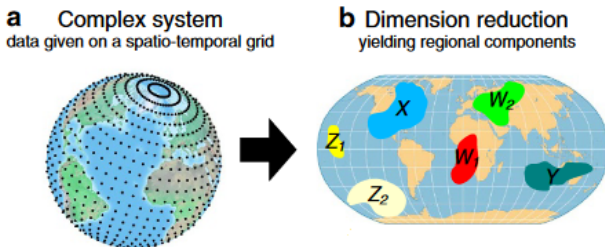
# Temperatura da Superfície Oceânica

- ▶ SST: Sea Surface Temperature;
- ▶ A temperatura da superfície oceânica é um dos fatores chave da mudança climática oceânica;
- ▶ Extremamente importante na compreensão do sistema climático terrestre.



# Temperatura da Superfície Oceânica

- ▶ Em estudos climáticos, é muito comum o estudo de modos de variância, os quais permitem a transformação de variáveis climáticas disponibilizadas em malhas temporais em duas componentes: uma de tempo e outra de espaço;
- ▶ São obtidas utilizando PCA com rotação Varimax.



**Figure:** Ilustração da obtenção de modos de variância a partir de dados espaçotemporais dispostos malha. Artigo: *Identifying causal gateways and mediators in complex spatio-temporal systems*, J. Runge.

# Temperatura da Superfície Oceânica

- ▶ Exemplo considerando-se um conjunto de dados de pressão superficial atmosférica;

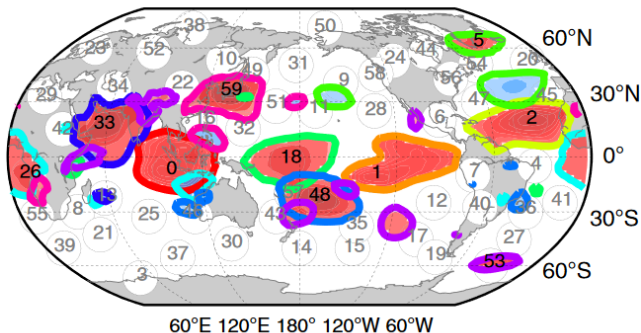
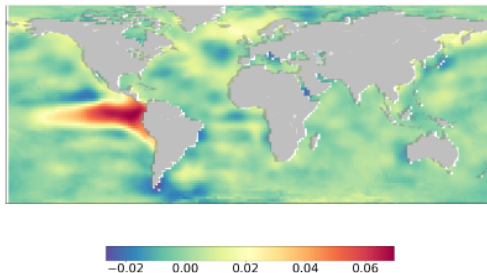


Figure: Artigo: *Identifying causal gateways and mediators in complex spatio-temporal systems*, J. Runge.

# Temperatura da Superfície Oceânica

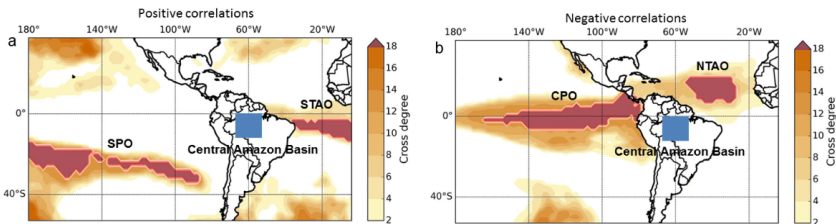
- ▶ A mudança da SST está relacionada aos mais diversos fenômenos climáticos;
- ▶ Um exemplo conhecido são as estiagens na Etiópia, as quais sabemos que tem forte relação causal com a SST no oceano Pacífico (ENSO). Evento de La Niña possui relação com estiagens no país.



**Figure:** Exemplo de modo de variancia com efeito no regime de chuvas etíope. Artigo: *Discovering Causal Factors of Drought in Ethiopia*, Mohammad Noorbakhsh

# Temperatura da Superfície Oceânica

- ▶ Outros estudos, utilizam de correlações cruzadas entre as séries temporais de SST para encontrar subprocessos relevantes à ocorrência de um fenômeno.



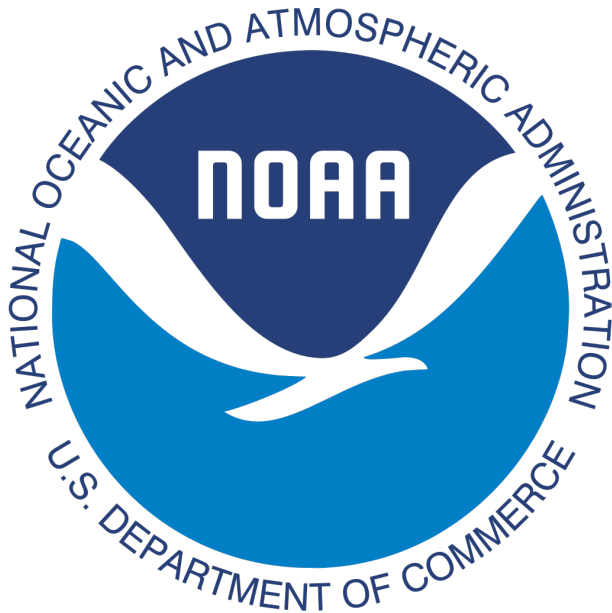
**Figure:** Correlação cruzada entre SST e anomalias hidrológicas continentais. Artigo: *An early-warning indicator for Amazon droughts exclusively based on tropical Atlantic sea surface temperatures*, Catrin Ciemer

# Temperatura da Superfície Oceânica

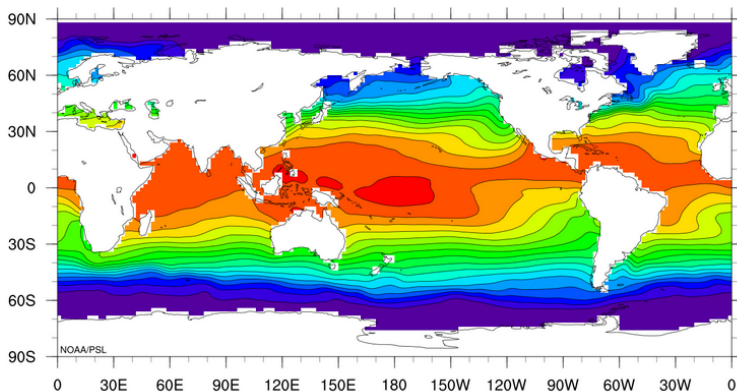
- ▶ Prever a SST é de grande interesse pois ela é uma variável causalmente relacionada ao comportamento de fenômenos climáticos e os subprocessos envolvidos;
- ▶ Porém, sua previsão é um desafio.
- ▶ Dados de SST são conhecidos por possuir informação redundante;
- ▶ Comportamento de variáveis climáticas é extremamente complexo considerando o sistema em que está inserido;
- ▶ Sazonalidade e padrões emergentes;
- ▶ Aqui, propomos um estudo de diferentes métodos de previsão da literatura com o objetivo de comparar performance.



# Dados de SST disponibilizados pela NOAA



- ▶ Para o nosso estudo, utilizamos os dados de reconstrução estendida de SST da NOAA (Physical Science Lab) em sua versão 5;
- ▶ Dados temporais em malha regular de SST compreendidos pela média mensal das temperaturas em degC.



**Figure:** Dados reconstruídos da média mensal da SST, novembro de 2023  
.NOAA Extended Reconstructed SST V5.

# Conjunto de dados

- ▶ Formato netCDF;
- ▶ Séries históricas 1854/01- 2023/11;
- ▶ 16.020 séries temporais;
- ▶ 32.648.760 observações;
- ▶ Atualizado mensalmente.

# Células de Circulação

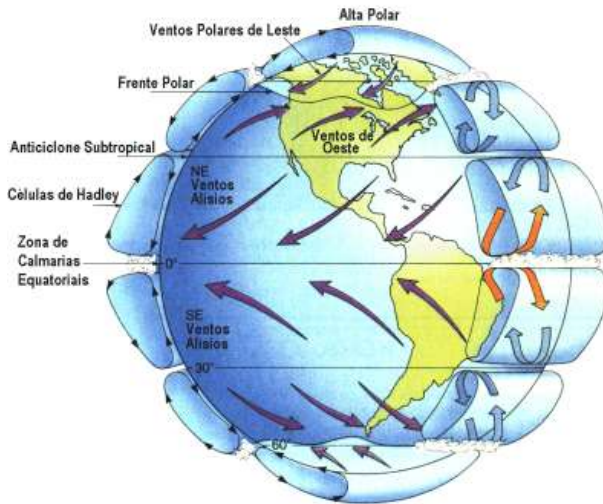
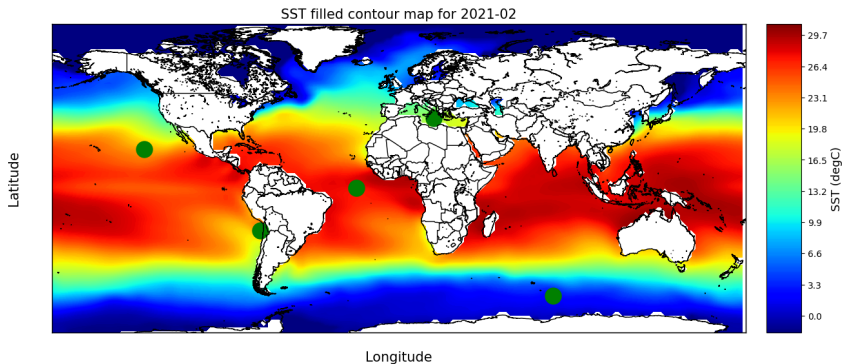


Figure: Imagem obtida do blog da UFPR,  
<http://fisica.ufpr.br/grimm/aposmeteo/cap8/cap8-1.html>

# Pontos de estudo escolhidos

- ▶ Um ponto de borda da região do ENSO ( $\text{lat} = 20^\circ$ ,  $\text{lon} = -132^\circ$ );
- ▶ Um ponto na costa do Chile (Corrente de Humboldt) ( $\text{lat} = -22^\circ$ ,  $\text{lon} = -72^\circ$ );
- ▶ Um ponto no meio do Atlântico ( $\text{lat} = 0^\circ$ ,  $\text{lon} = -22^\circ$ );
- ▶ Um ponto no meio do Mediterrâneo ( $\text{lat} = 36^\circ$ ,  $\text{lon} = 18^\circ$ );
- ▶ Um ponto no extremo Índico (ponto de maior latitude) ( $\text{lat} = -56^\circ$ ,  $\text{lon} = 80^\circ$ ).

# MAPA: Pontos de estudo escolhidos



# O modelo SARIMA

- ▶ Os modelos Autoregressivos de Média Móvel, são usados para prever valores futuros das séries
- ▶ São compostos por duas partes: a parte autoregressiva (AR) e a parte de média móvel (MA).
- ▶ SARIMA inclui termos sazonais e não sazonais
- ▶ Para o problema usado Horizonte = 12 meses
- ▶ Hiperparâmetros order = (p, q, d) e sazonal\_order = (P, Q, D, M)

## O modelo LSTM

- ▶ Long Short-Term Memory
- ▶ Implementado com Pytorch

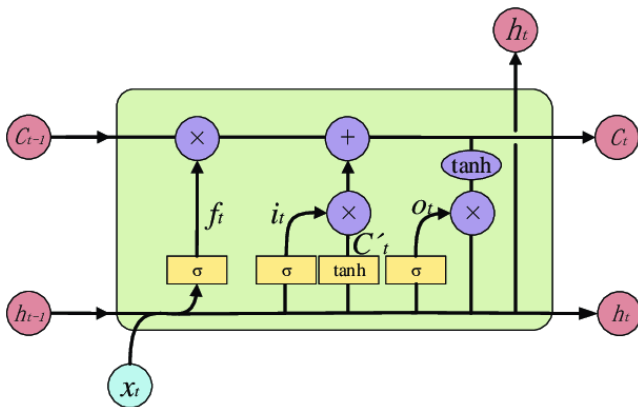


Figure: Arquitetura do LSTM. Fonte: <https://medium.com/analytics-vidhya/exploring-different-types-of-lstms-6109bcb037c4>



# O modelo LSTM

## Caratcterísticas de treinamento

- ▶ Lookback = 12 meses;
- ▶ 1 camada LSTM com 64 *features* no *hidden state*  $h$ ;
- ▶ 1 camada linear densa de entrada de tamanho 64 e saída de tamanho 1;
- ▶ Otimizador Adam,  $lr = 0.002$ ;
- ▶ Função perda MSELoss;
- ▶ 100 épocas de treino.

# O modelo LSTM

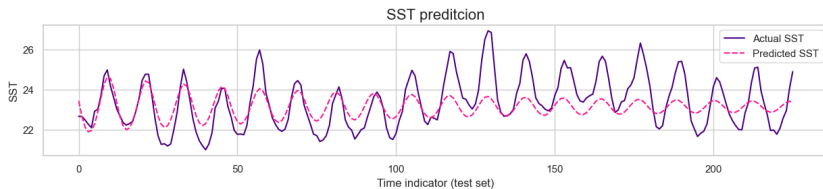


Figure: Treino com 1800 observações e teste em 226

# O modelo LSTM

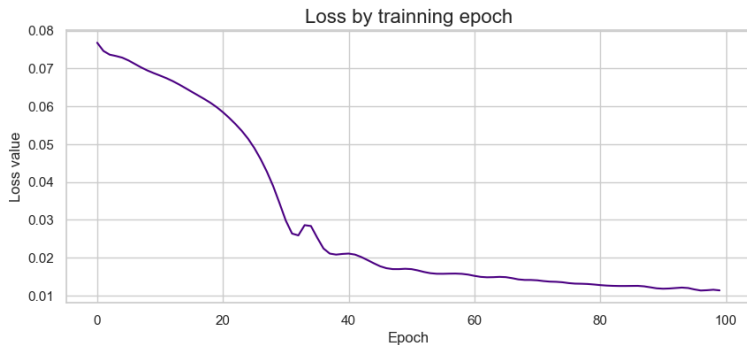


Figure: Perda no decorrer do treinamento

# O modelo DS

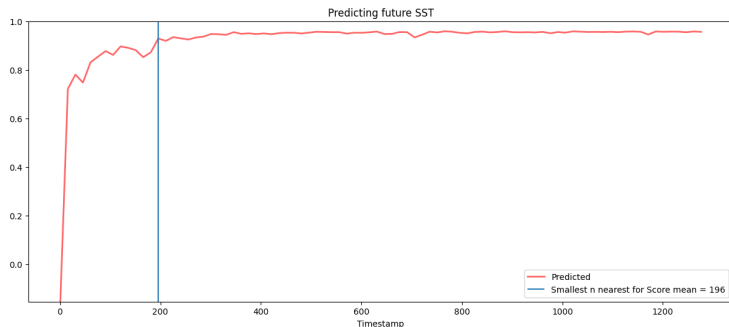
- ▶ Foi implementada a técnica de forecasting apresentada pelo artigo **Improved Particle Swarm Optimization for Sea Surface Temperature Prediction** nomeada DS. O nome vem da junção DTW + SVR.
- ▶ Para prever as próximas  $x$  observações utilizamos a janela referente as últimas observações que possuímos e procuramos subséries anteriores de mesmo tamanho utilizando DTW.
- ▶ As subséries mais próximas e as respectivas próximas  $x$  observações encontradas são utilizadas para treinar um SVR.
- ▶ Não consideramos a otimização de hiperparâmetros com meta-heurísticas

# O modelo DS

## Características de treinamento

- ▶ Series Sample Ratio: 100%
- ▶ Top Subseries Ratio: 30%
- ▶ Window Size: 96
- ▶ SVR Kernel: Radial Basis Function

# O modelo DS



**Figure:** Analisando precisão de acordo com a quantidade de subséries utilizadas no treinamento do SVR. Fonte: Autor

# O modelo DS

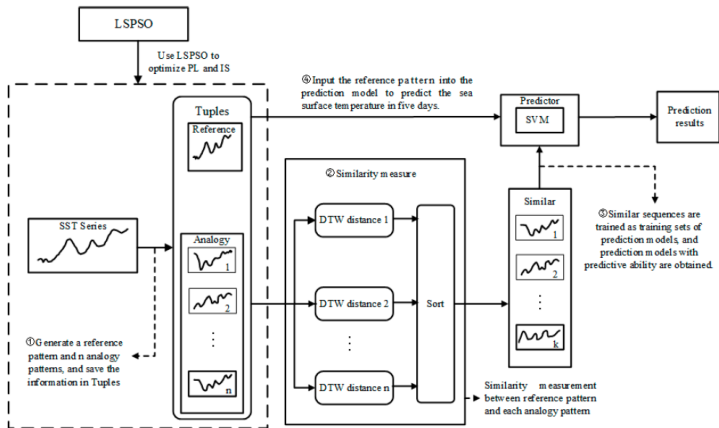
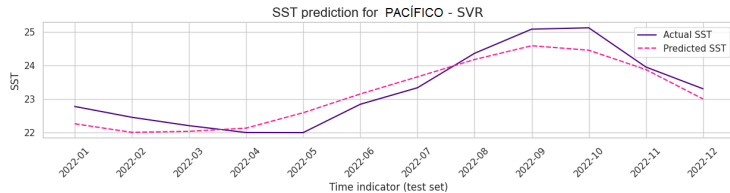
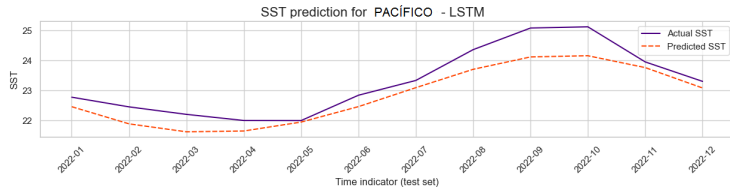
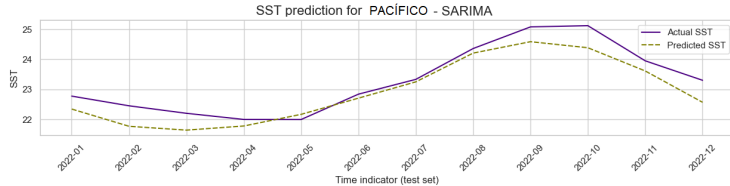


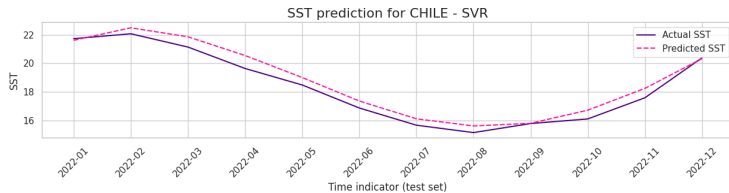
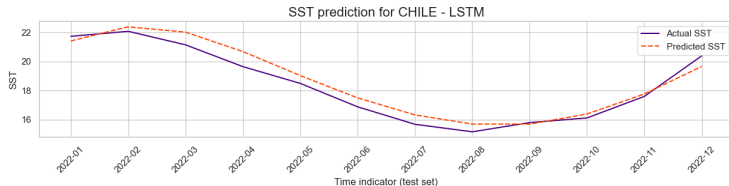
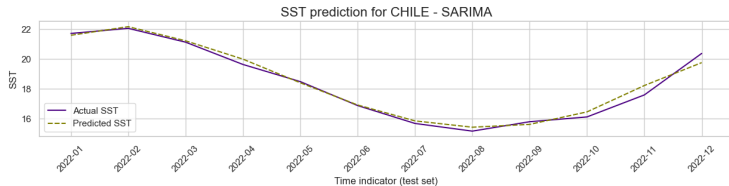
Figure: Arquitetura do modelo DS. Fonte:  
<https://www.mdpi.com/1996-1073/13/6/1369>

# Alguns resultados previsões (PACÍFICO)

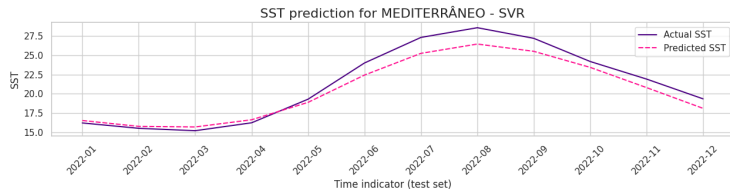
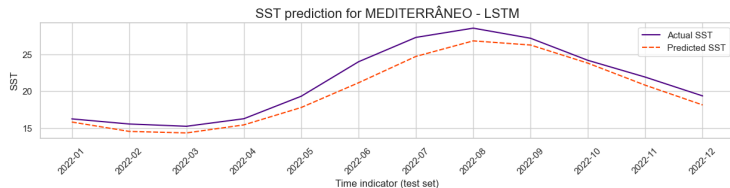
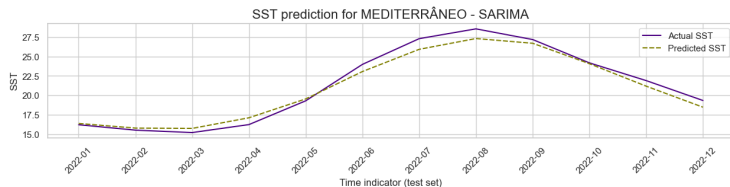




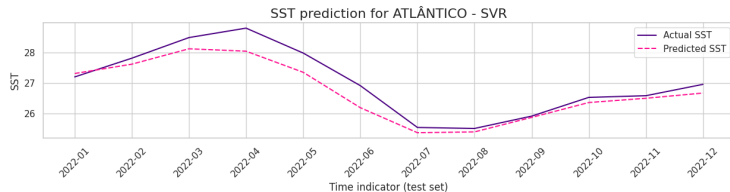
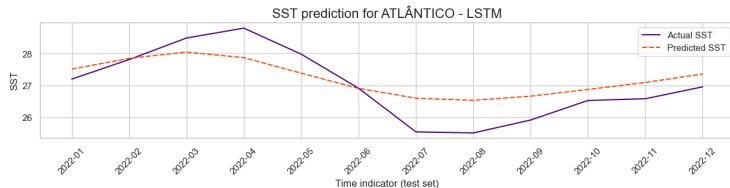
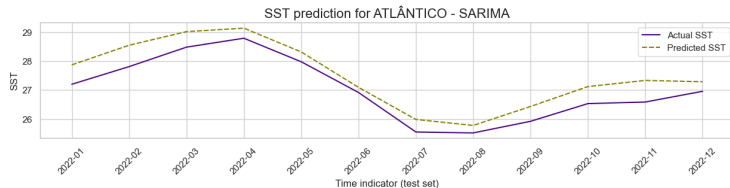
# Alguns resultados previsões (CHILE)



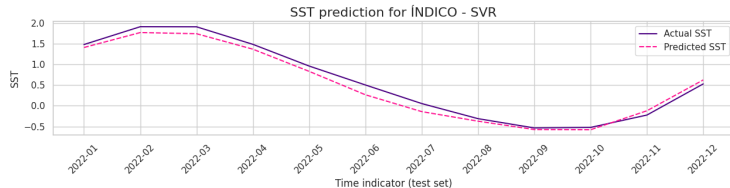
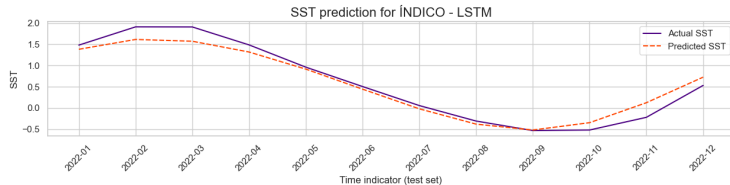
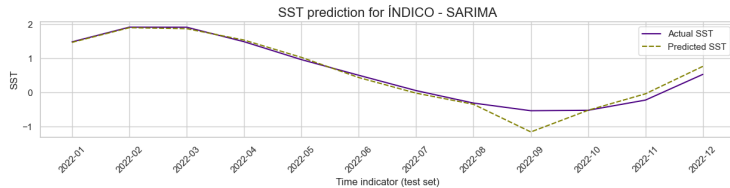
# Alguns resultados previsões (MEDITERRÂNEO)



# Alguns resultados previsões (ATLÂNTICO)



# Alguns resultados previsões (ÍNDICO)



# Métricas de avaliação: MAPE e RMSE

Erro absoluto percentual médio:  $MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$

Raiz do erro quadrático médio:  $RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2}$

# Tabela comparativa: RMSE e MAPE

Tabela comparativa: RMSE e MAPE						
Região	SARIMA		LSTM		DS	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
ENSO	0.45	0.016	0.54	0.019	<b>0.396*</b>	<b>0.015**</b>
Chile	<b>0.31*</b>	<b>0.014**</b>	0.58	0.028	0.521	0.025
Atlântico	0.76	0.03	0.63	0.02	<b>0.391*</b>	<b>0.011**</b>
Mediterrâneo	<b>0.50*</b>	<b>0.02**</b>	1.48	0.06	1.224	0.044
Índico	0.20	<b>0.34**</b>	0.19	0.38	<b>0.130*</b>	0.456

\* menor valor RMSE para o modelo e região

\*\* menor valor MAPE para o modelo e região

# SARIMAX: Walk-forward validation

- ▶ Previsões ao longo do tempo tornam-se cada vez menos precisas. Uma abordagem mais realista seria treinar novamente o modelo com dados reais à medida que ficam disponíveis.
- ▶ A solução walk-forward é usada para obter resultados mais precisos
- ▶ É uma alternativa ao cross-validation para modelos de séries temporais

# Walk-foward validation - Ideia

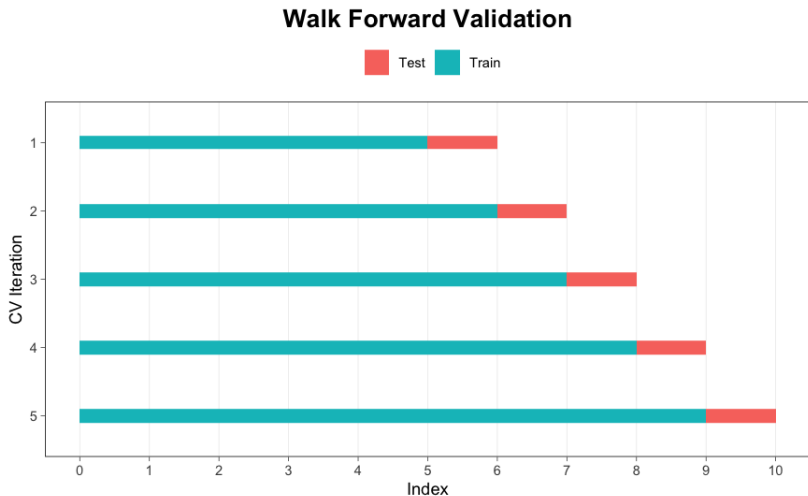
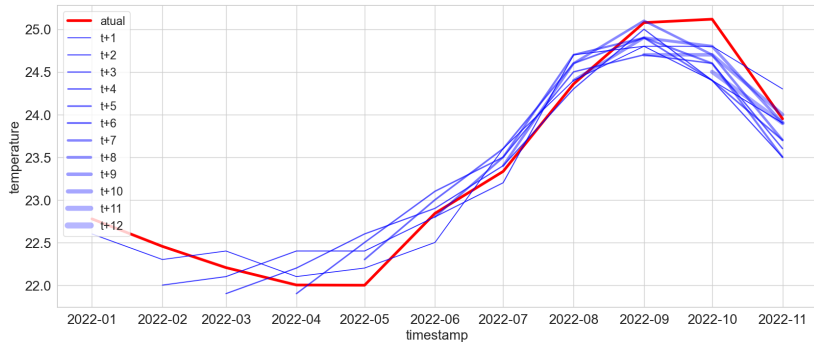


Image credits: <https://github.com/AaronOSO/Time-series-Cross-validation>



# Resultado: previsões usando walk-foward (ENSO)



# Conclusão

- ▶ O DS consegue aproximar modelos conhecidos e amplamente utilizados em forecasting de séries temporais;
- ▶ Devido ao pequeno número de séries utilizadas, principalmente pela dificuldade de paralelizar e analisar grande quantidade de testes não é possível dizer que houve uma melhora significativa na previsão;
- ▶ Entretanto, o modelo proposto tem potencial e é uma solução tangível em séries que apresentam periodicidade;
- ▶ A técnica permite o treino de um modelo com menos dados redundantes e com potencial de maior robustez.

# Próximos Passos

- ▶ Otimizar os parâmetros do DS
- ▶ Ponderar subséries durante treinamento do SVR
- ▶ Quantificar regularidade da série antes da modelagem, utilizando medidas como Approximate Entropy
- ▶ Amostrar subséries ao invés de utilizar DTW por toda extensão e estudar a convergência
- ▶ Expandir e automatizar testes