

# Supplementary Material

## Predicting Atlantic and Benguela Niño events with deep learning

**Marie-Lou Bachèlery<sup>1,2</sup>, Julien Brajard<sup>3</sup>, Massimiliano Patacchiola<sup>4</sup>, Serena Illig<sup>5,6</sup>, Noel Keenlyside<sup>1</sup>**

<sup>1</sup> Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway

<sup>2</sup> Climate Simulation and Prediction Division, Euro-Mediterranean Center on Climate Change, Bologna, Italy

<sup>3</sup> Nansen Environmental and Remote Sensing Center, Bergen, Norway

<sup>4</sup> Department of Engineering, University of Cambridge, Cambridge, United Kingdom

<sup>5</sup> Laboratoire d'Études en Géophysique et Océanographie Spatiale (LEGOS), CNRS/IRD/UT3/ CNES, Toulouse, France.

<sup>6</sup> Department of Oceanography, University of cape town, cape town, South Africa.

### **Including:**

Supplementary Figure A

Supplementary Figure B

Supplementary Figure C

Supplementary Figure D

Supplementary Figure E

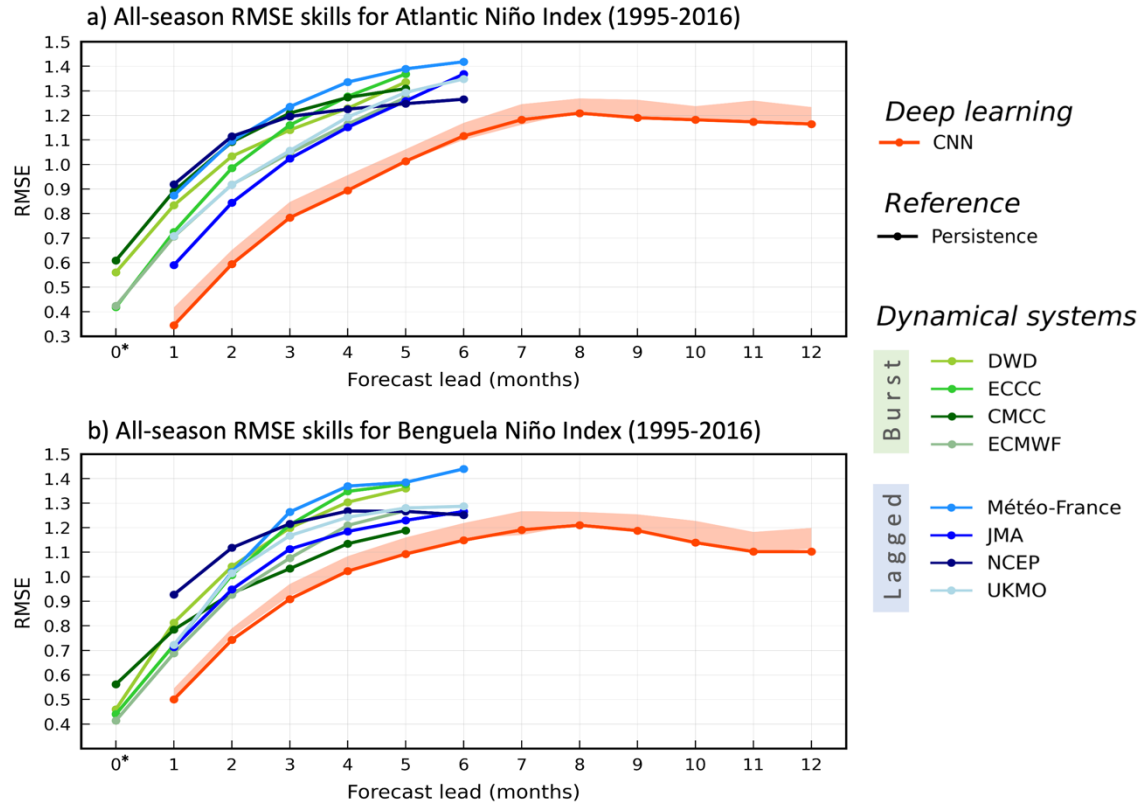
Supplementary Figure F

Supplementary Figure G

## Introduction:

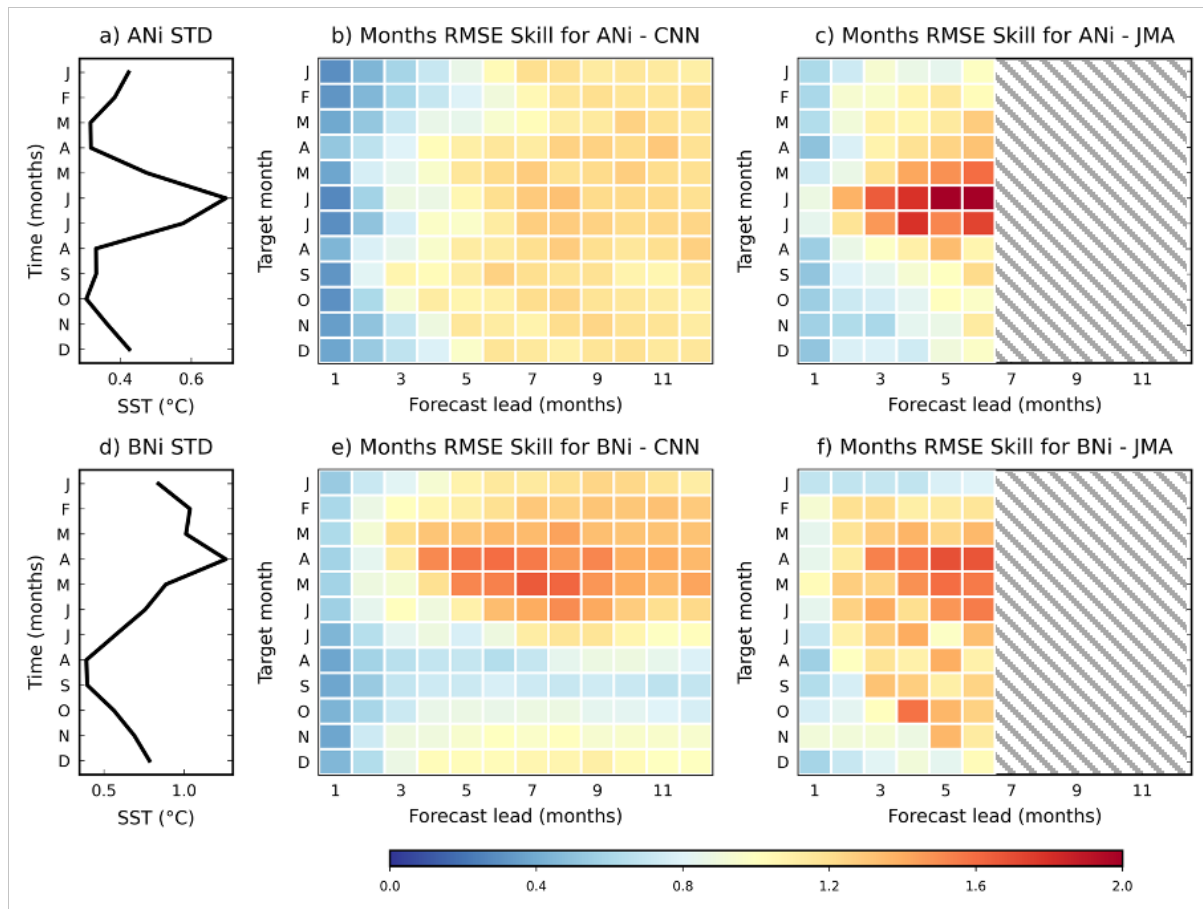
These **supplementary materials** provide extra diagnostics to assess the prediction skill of the CNN model compared with the C3S seasonal prediction systems.

### 1 Supplementary Note 1: All-months Root Mean Square Error (RMSE) skills assessment

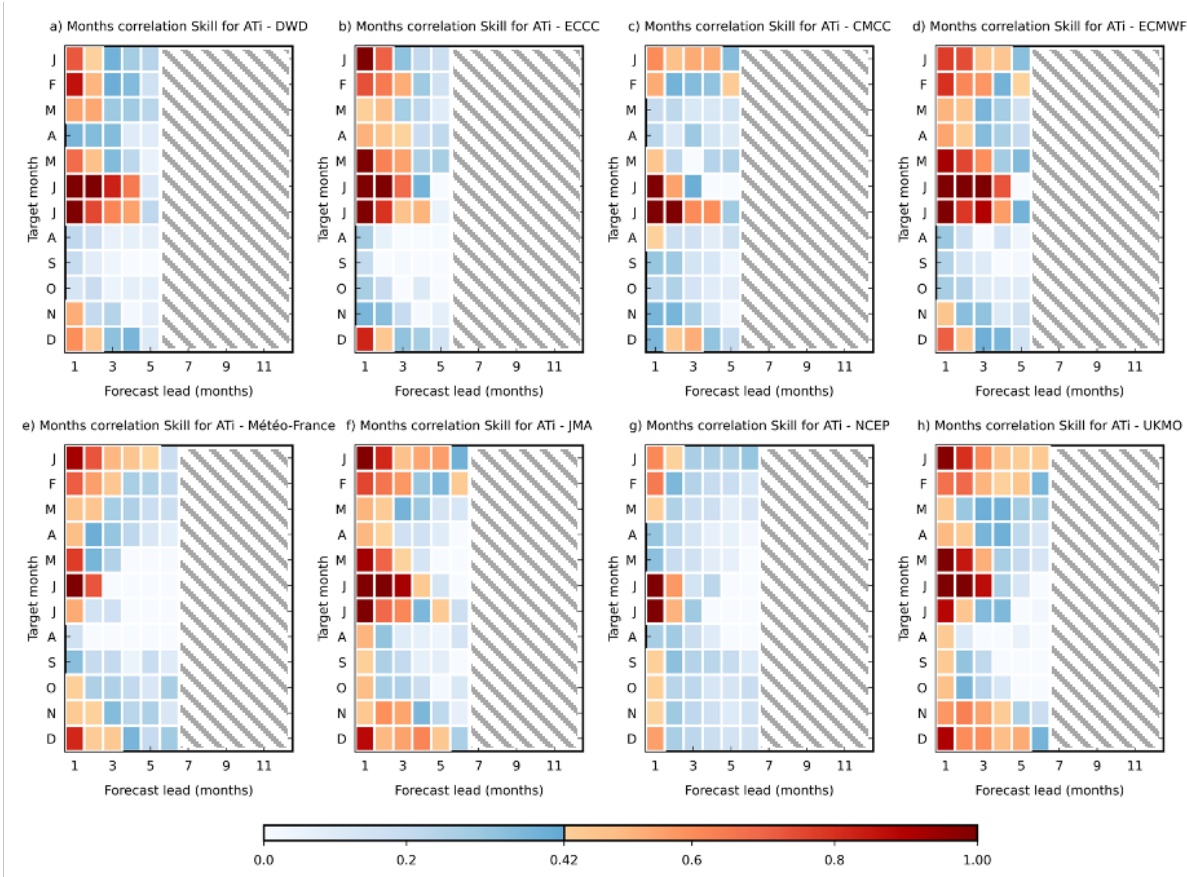


**Figure A: Root Mean Square Error (RMSE) performances of state-of-the-art C3S dynamical systems and our deep learning model.** (a) The Atlantic Niño index [ANi: 20°W-0°E/3°S-3°N] correlation as a function of lead time (months) for the CNN deep-learning model (orange), the “burst” (green colors), and the “lagged” C3S prediction models (blue colors). The orange shading shows the highest and the lowest correlations estimated over 20 ensemble members of the CNN model. Persistence is shown in black and the statistical significance threshold (>99%, estimated using a bootstrap method<sup>58</sup> (see Methods) is indicated by the black horizontal line. The evaluation period is 1995-2016 and all models are compared with ORAS5. (c) Same as panel b) for the Benguela Niño index [BNi: 10°S-20°S/2° coastal band].

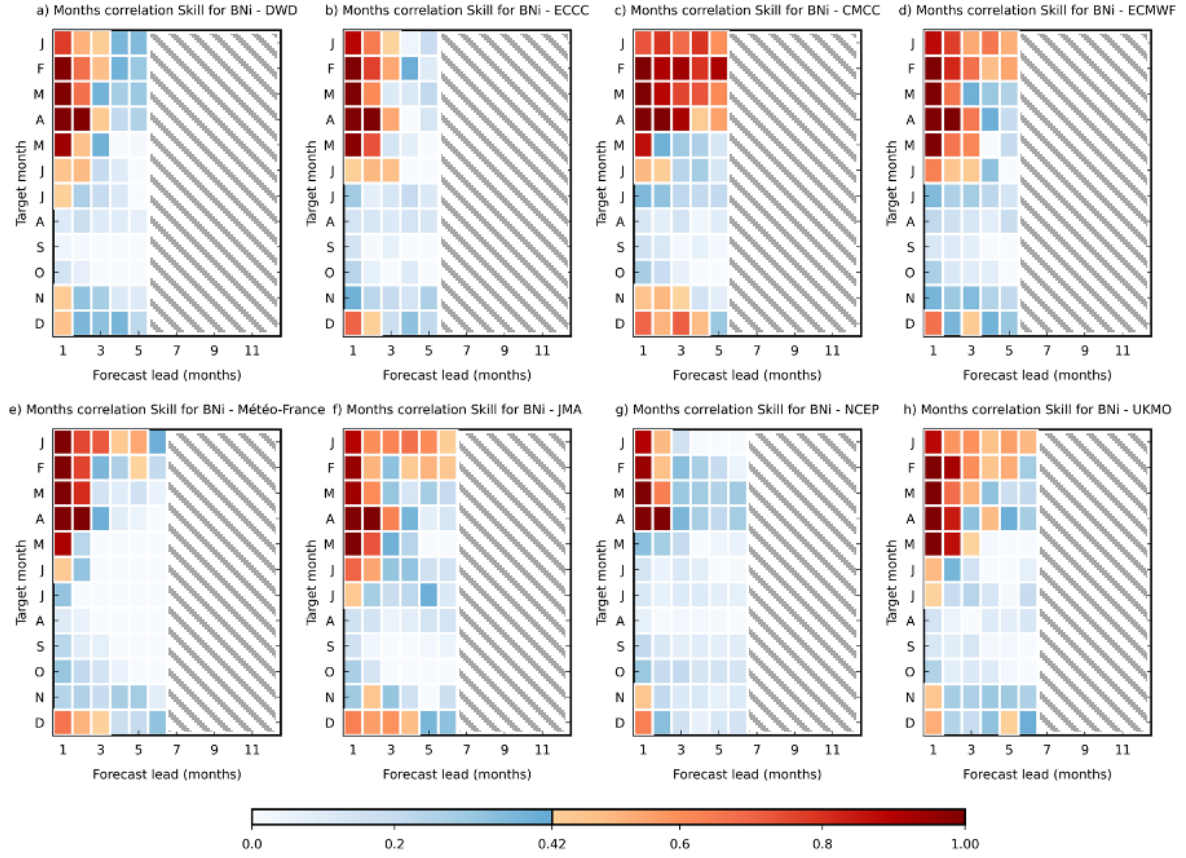
## 2 Supplementary Note 2: Seasonal decomposition skills of the CNN and C3S models



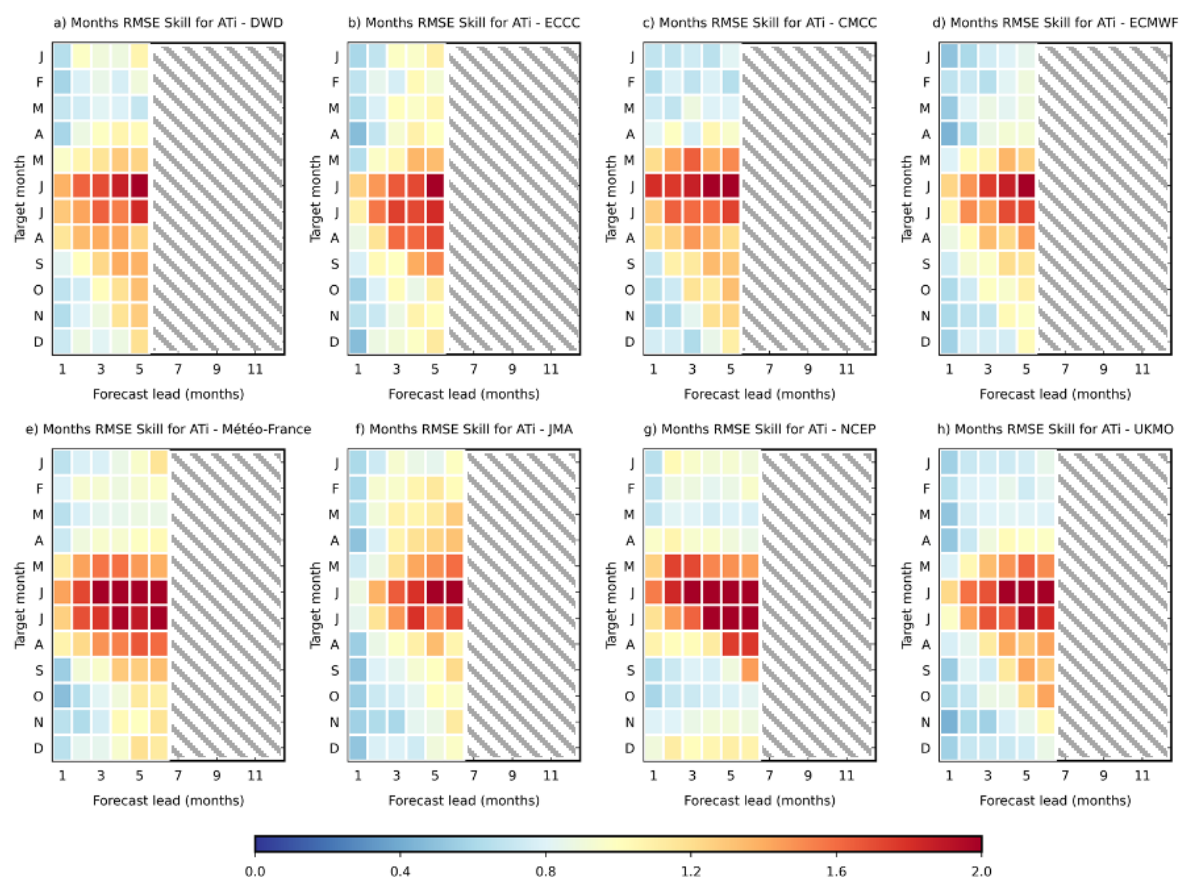
**Figure B: Seasonal Root Mean Square Error (RMSE) decomposition skills** for the Atlantic Niño (top panels) and Benguela Niño (bottom panels) indexes. (left) Monthly standard deviation of ORAS5 (°C). Seasonal rmse decomposition (see Methods) as a function of the lead time (months) for the CNN deep-learning model (middle) and the JMA dynamical model (right). The evaluation period is 1995-2016 and all models are compared to ORAS5. Statistically significant values (>99%, estimated using a bootstrap method<sup>58</sup>; see Methods) are indicated by the red color bar.



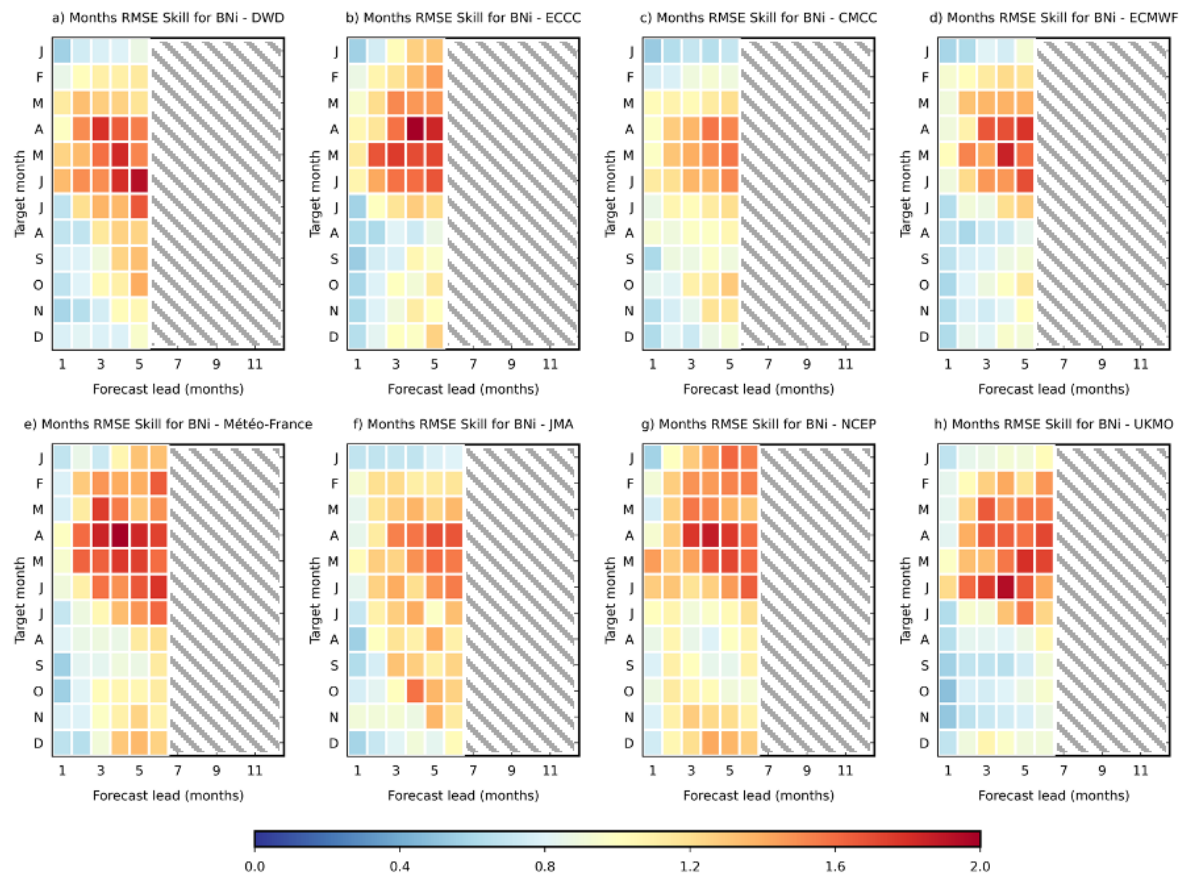
**Figure C: Seasonal correlation decomposition skills of C3S for the ANi.** Seasonal correlation decomposition (see Methods) as a function of the lead time (months) for a) DWD, b) ECCC, c) CMCC, d) ECMWF, e) Météo-France, f) JMA, g) NCEP and UKMO. The evaluation period is 1995-2016 and all models are compared to ORAS5. Statistically significant values (>99%, estimated using a bootstrap method<sup>58</sup>; see Methods) are indicated by the red color bar.



**Figure D: Seasonal correlation decomposition skills of C3S for the BNi.** Same as Fig. C for the BNi.

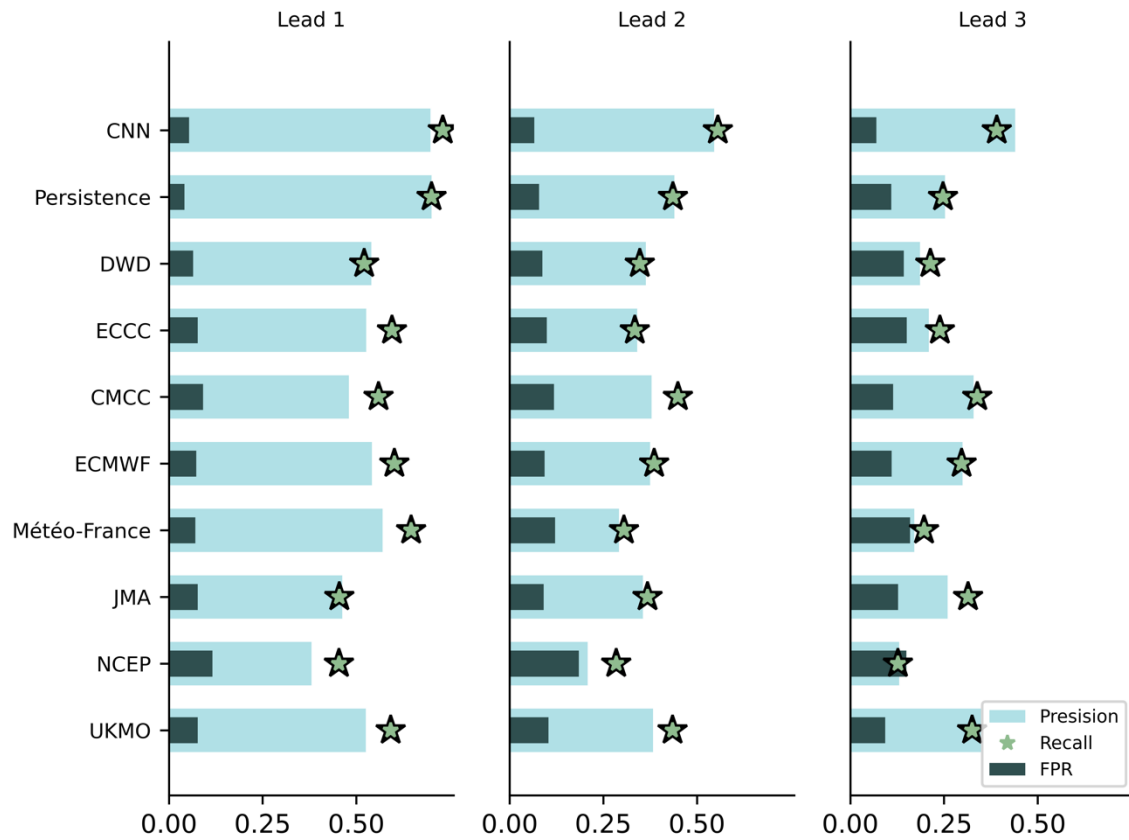


**Figure E: Seasonal root mean square error decomposition skills of C3S for the ANi.** Seasonal rmse decomposition (see Methods) as a function of the lead time (months) for a) DWD, b) ECCC, c) CMCC, d) ECMWF, e) Météo-France, f) JMA, g) NCEP and UKMO. The evaluation period is 1995-2016 and all models are compared to ORAS5.



**Figure F: Seasonal root mean square error decomposition skills of C3S for the BNi.** Same as Fig. E but for BNi.

### 3 Supplementary Note 3: Benguela Niño/Niña events detection skills



**Figure G: Ability to predict the occurrence of extreme Benguela Niños/Niñas.** Precision (light blue bars), recall (stars) and False Positive Rate (FPR; dark green bars; see Methods) for Benguela Niño/Niña events combined, using the CNN model, the persistence (ORAS5), and the C3S dynamical prediction systems. Scores are evaluated against ORAS5 for leads 1 (left panel) to 3 (right panel). Events are selected when the BNi amplitude exceeds 1 standard deviation for two consecutive months.