cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

AI-enhanced seasonal predictions of Mediterranean cyclones

Leone Cavicchia

S. Gualdi, E. Scoccimarro (CMCC)

G. Ascenso, M. Giuliani, A. Castelletti (PoliMI)

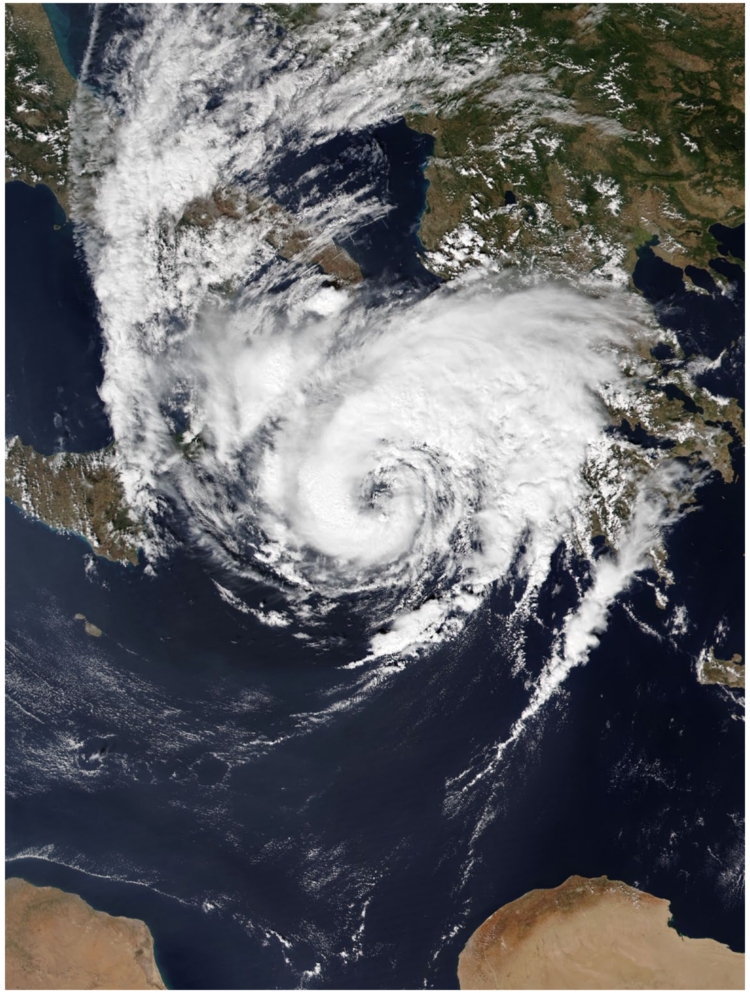
ATMOS 2024, Bologna

4-7-2024

www.cmcc.it

Background

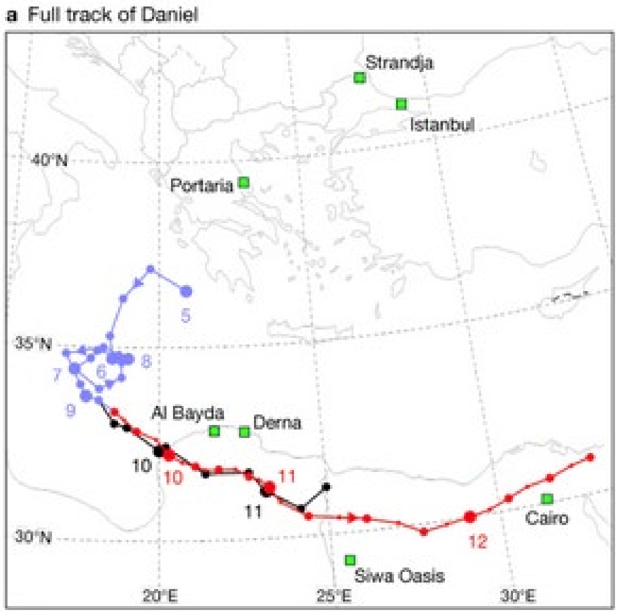
- Cyclones form frequently in the Mediterranean basin due to region location and the complex topography
- Even tough smaller and shorter lived than cyclones forming in other ocean basins Med cyclones cause severe damage in the highly populated coasts of the region
- A number of different dynamical mechanisms for cyclone genesis and intensification play a role, resulting in the occurrence of different types of low pressure systems, ranging from mid-latitude to tropical-like cyclones

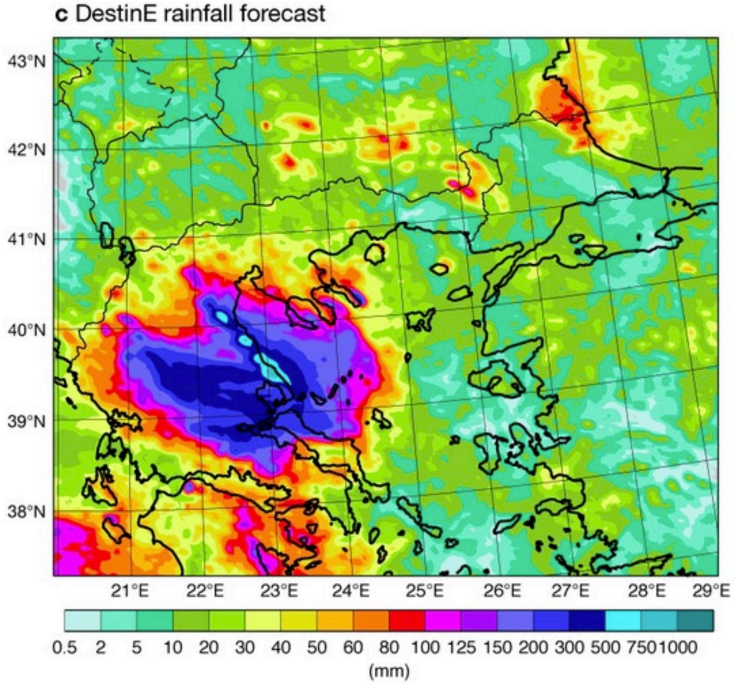


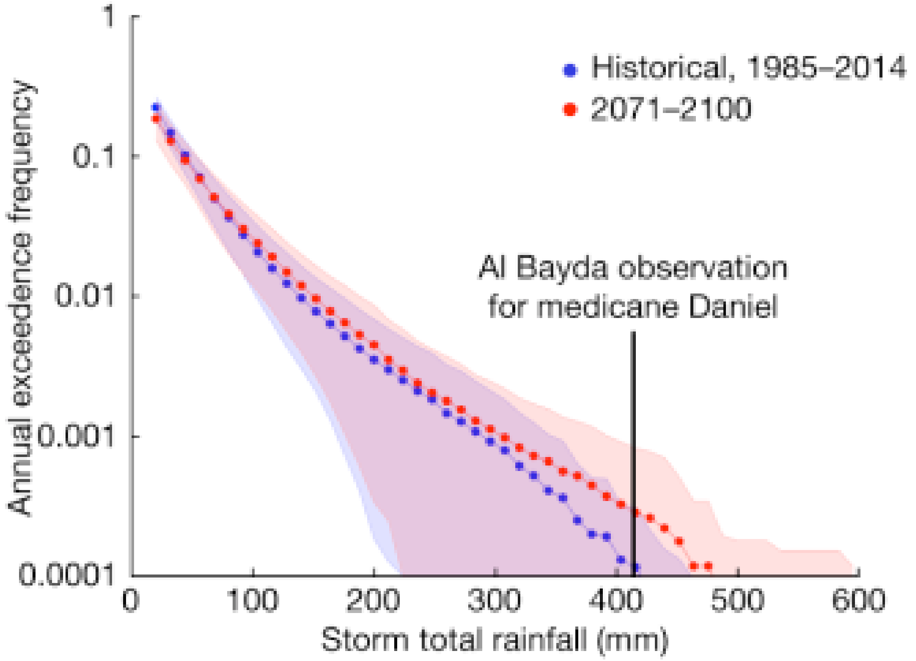
Ianos (September 2020)

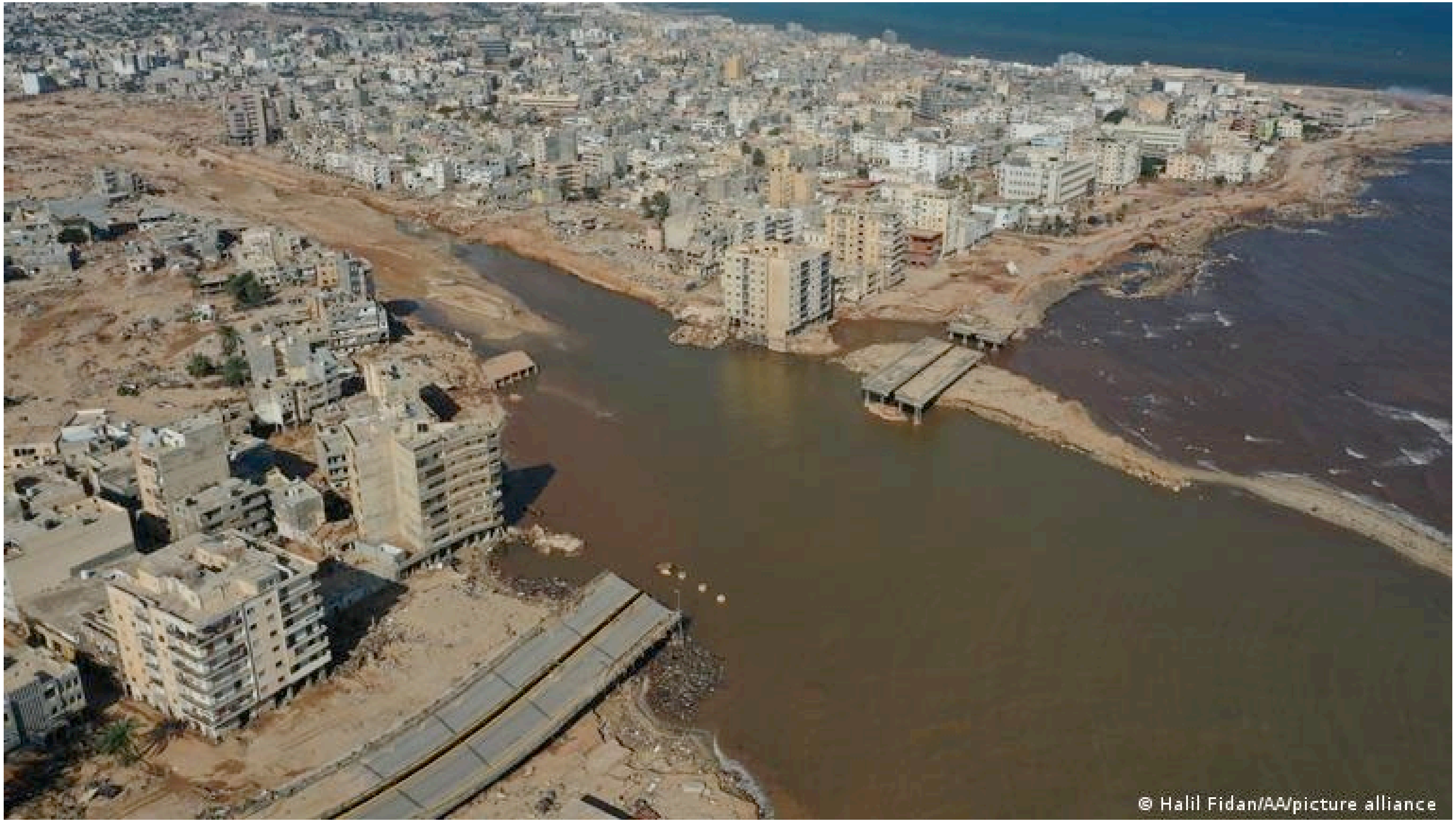
Storm Daniel

- *Storm Daniel in September 2023 was the costliest cyclone outside the North Atlantic (> 20 B US$)*
- *The deadliest cyclone globally since 2013 (10.000 fatalities estimated)*

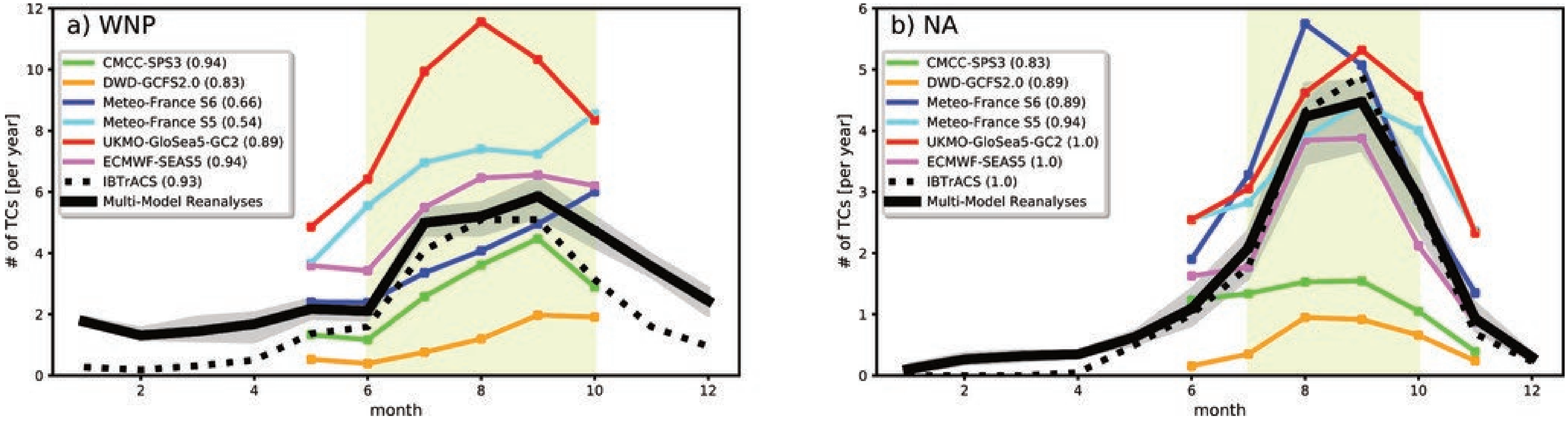






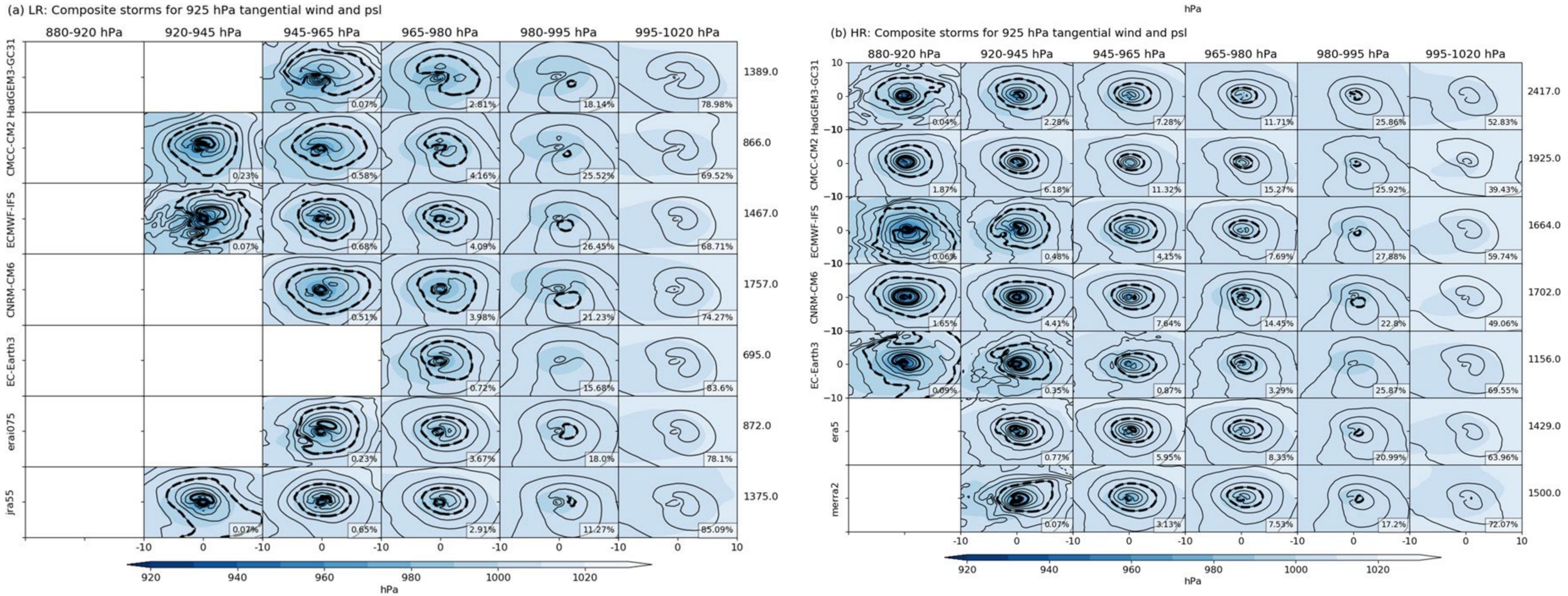

© Halil Fidan/AA/picture alliance

Cyclones in Seasonal forecast



Befort et al 2022

Climate model resolution and TCs

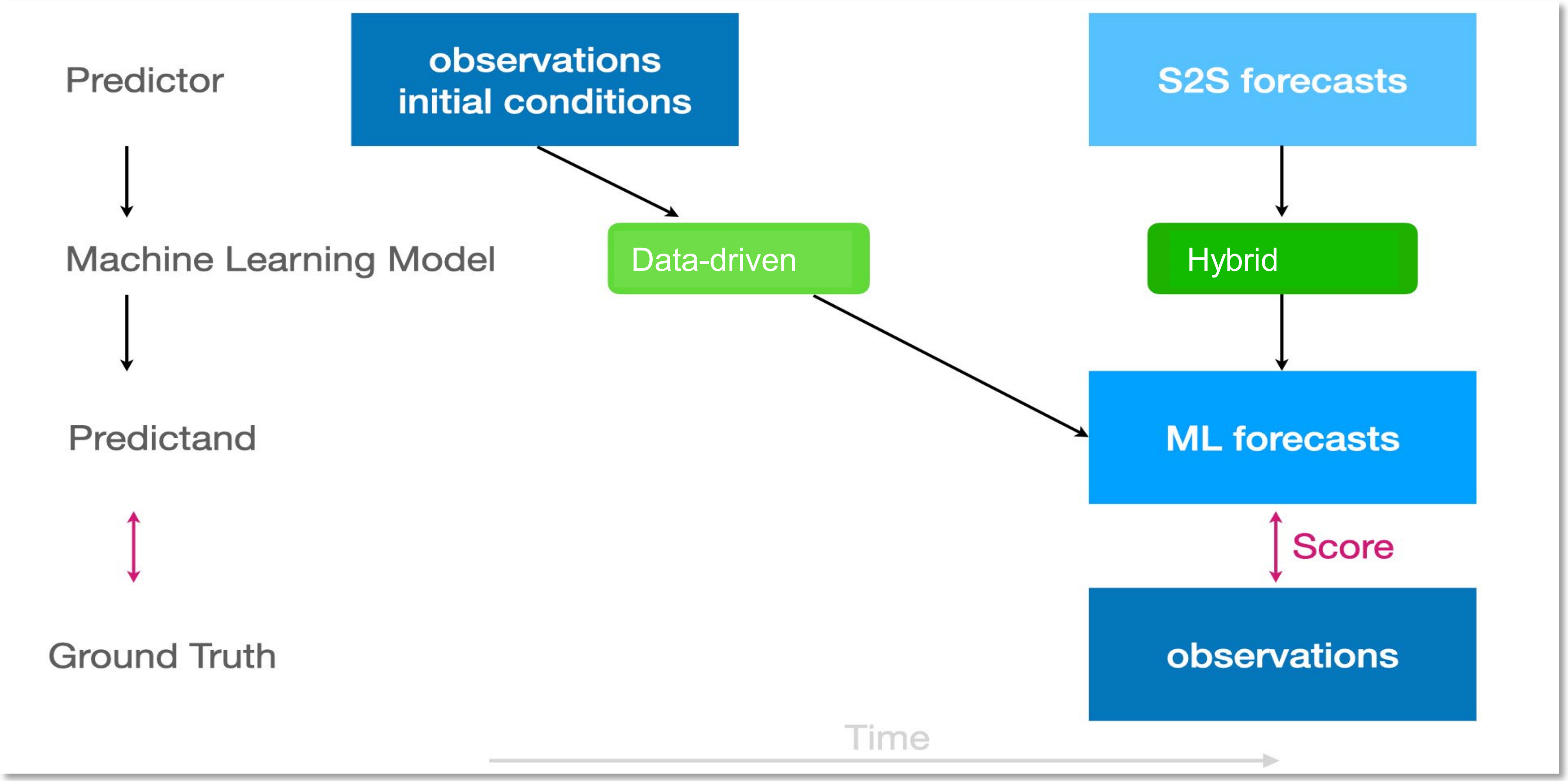


Roberts et al (2020): Impact of Model Resolution on Tropical Cyclone Simulation Using the HighResMIP–PRIMAVERA Multimodel Ensemble

Low-resolution models reproduce only a fraction of the observed cyclones, and are not able to reproduce intense cyclones

Two approaches to AI seasonal forecast

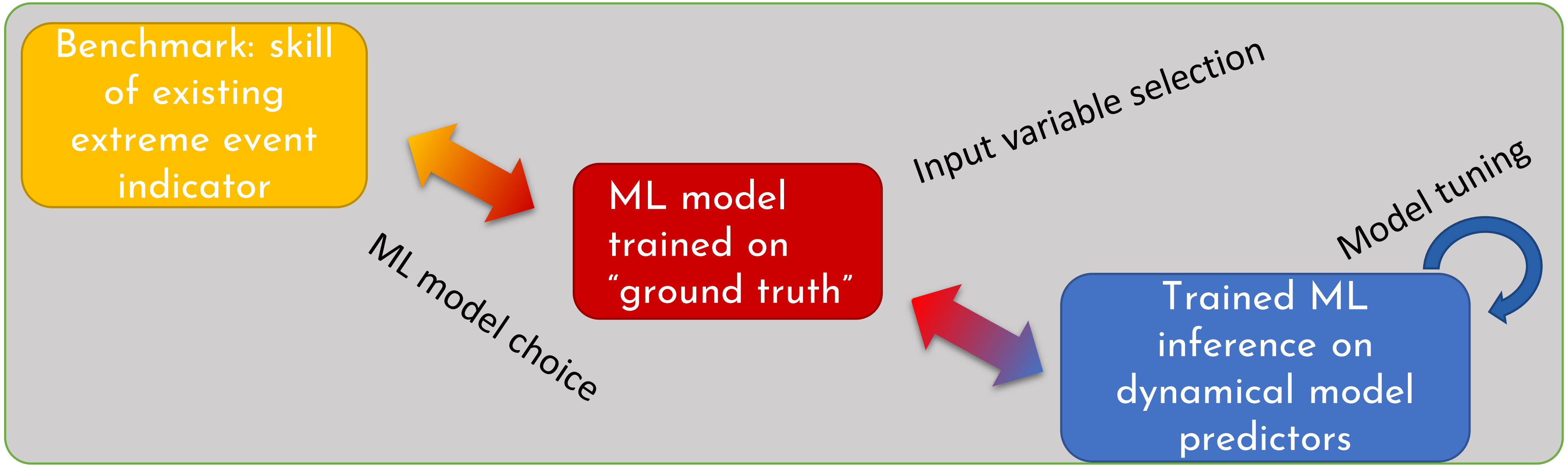
Adapted from : https://s2s-ai-challenge.github.io/



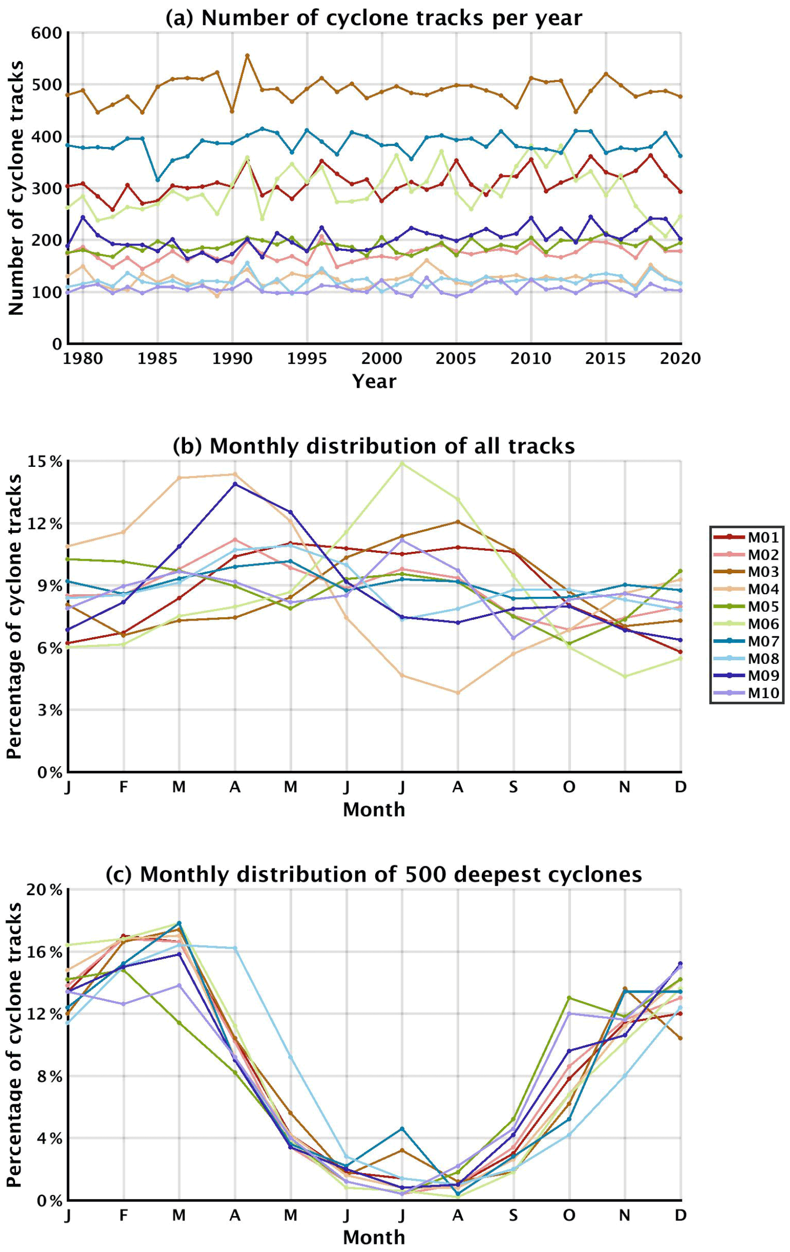
CYCLOPS:
AI-enhanced seasonal prediction of Mediterranean cyclones

The aim of this project is to improve the prediction of cyclone activity, exploiting ***a hybrid AI approach*** *where the occurrence of extremes is linked to large-scale meteorological fields produced by a dynamical model:*

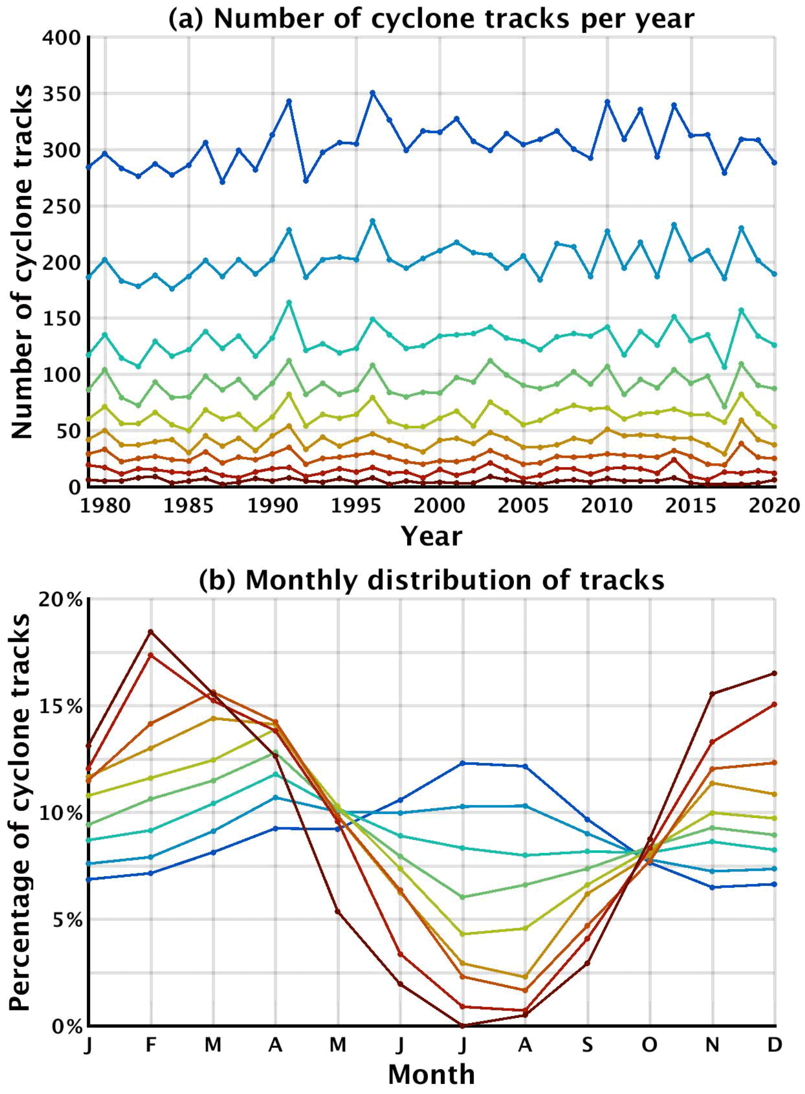
- *First the (statistical) connection between the large-scale variables (predictors) and the extreme of interest (predictand) is established in the "ground truth" by training one or several ML models on observational/reanalysis dataset*
- *The trained ML model is then applied in inference mode on the same large-scale predictors from the dynamical seasonal forecast model hindcasts, and the prediction compared with observations.*
- *The ML model is tuned to compensate the effect of the dynamical model bias on the predictive skill.*



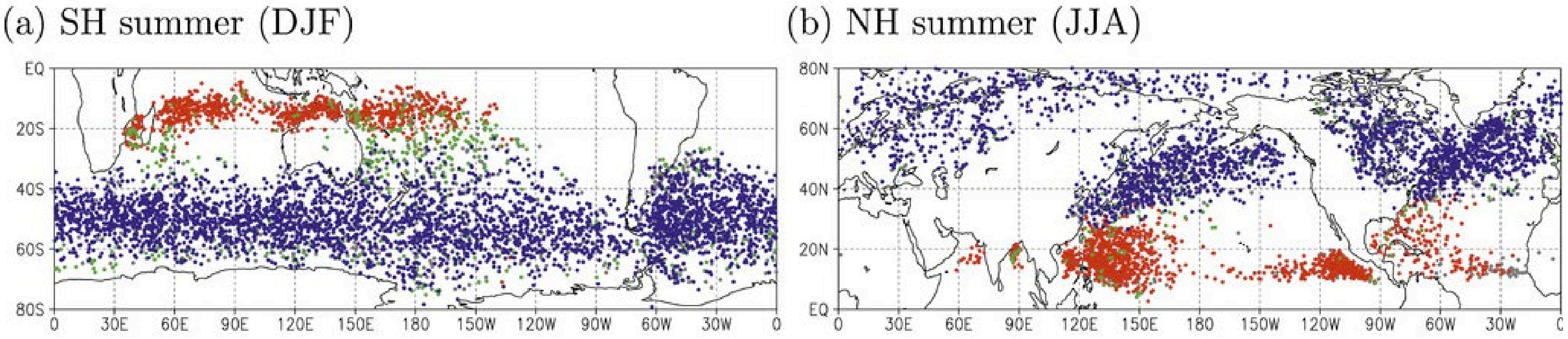
Data



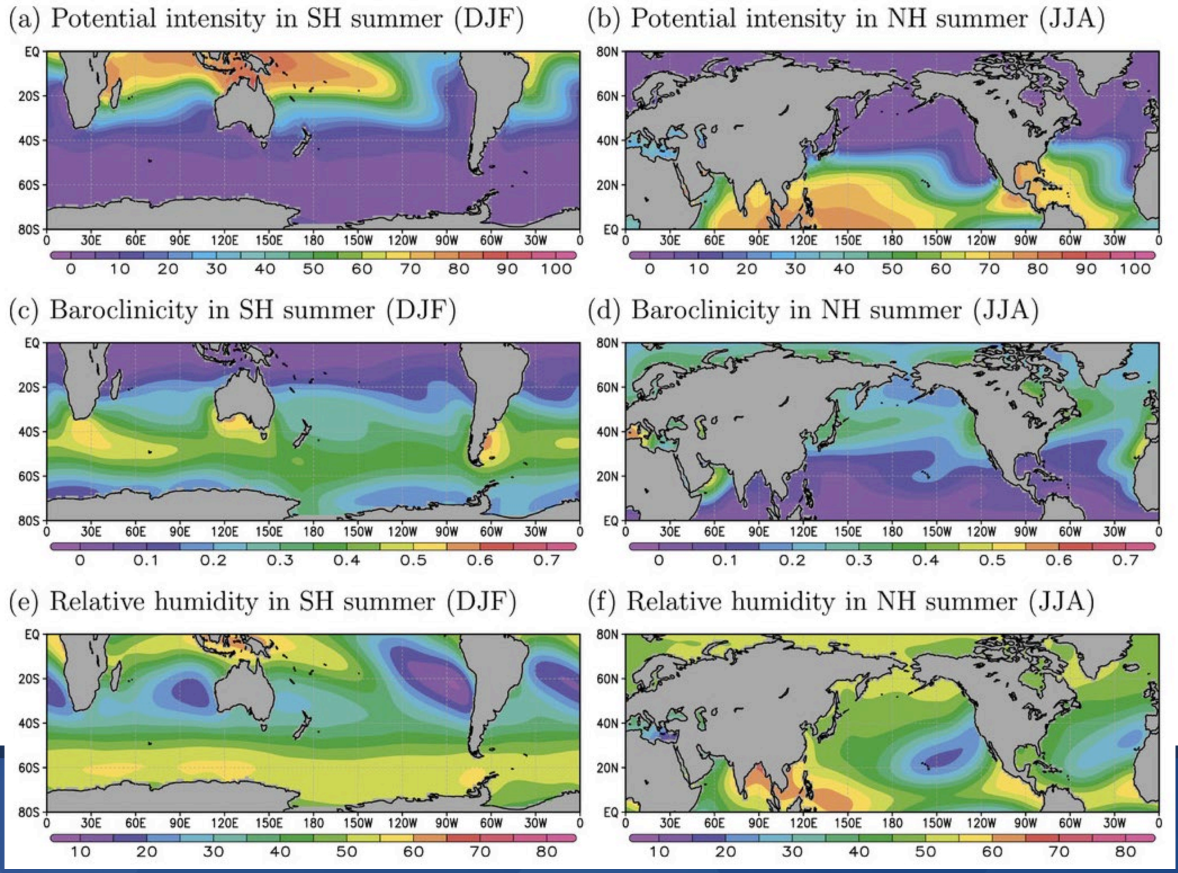
- No fully observation-based database (such as IBTrACS) available in the region
- Data from Flaounas et al. 2023 “best track” dataset, based on the consensus between ten different cyclone tracking algorithms applied to ERA5.
- In this work a confidence level of 7 has been used.



Drivers



From Yanase et al. 2014



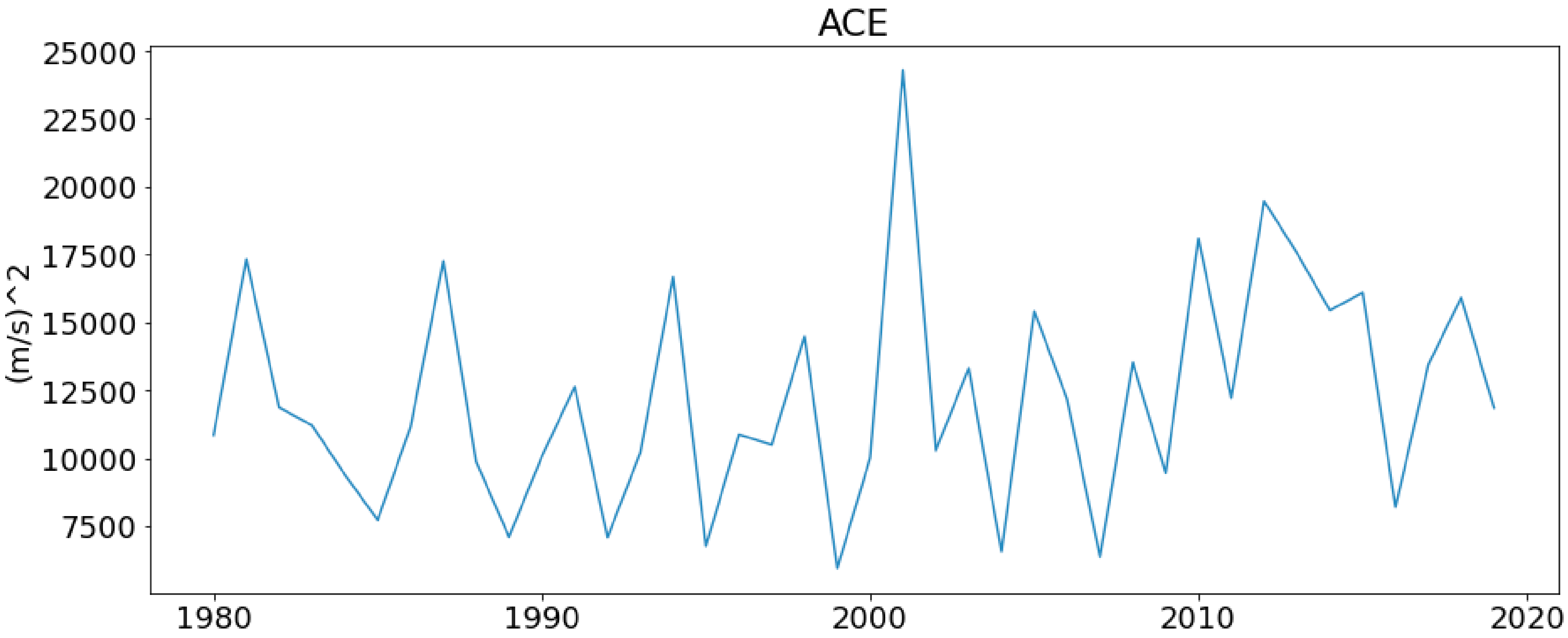
Predictors	
Tropical cyclogenesis	Vorticity Wind shear Humidity SST
Extratropical cyclogenesis	Eady growth rate

Target

Common metrics used for cyclone activity include cyclone number, cyclone days and ACE (accumulated cyclone activity).

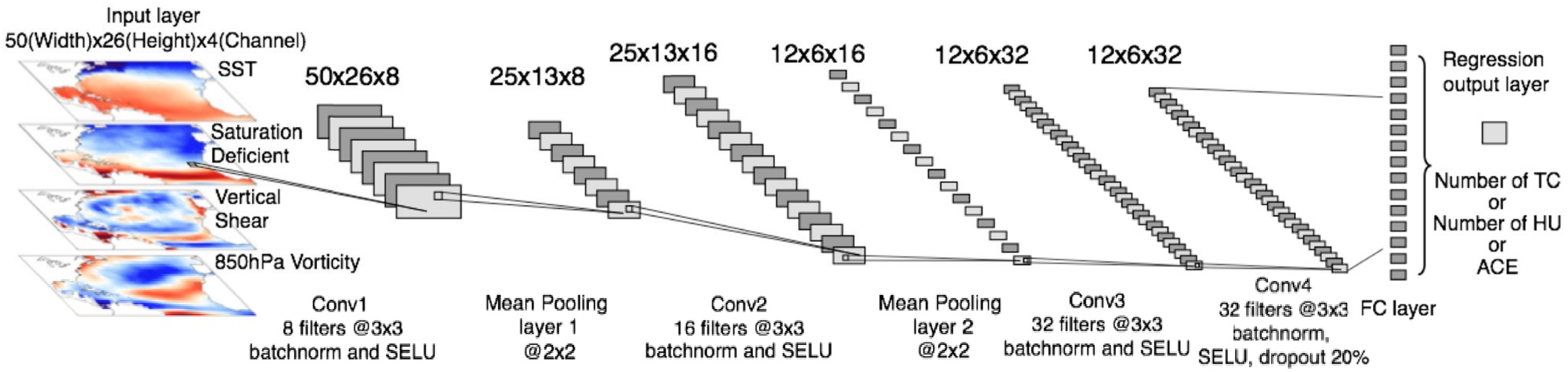
Here we focus on ACE, which has a number of advantages:

- Naturally gives more weight to more intense cyclones, with no need to impose ad hoc filters
- Less sensitivity on the details of the cyclone detection scheme used to produce the ground truth database



$$\mathrm{ACE} = 10^{-4} \sum v_{\max}^2$$

Model 1 (CNN)

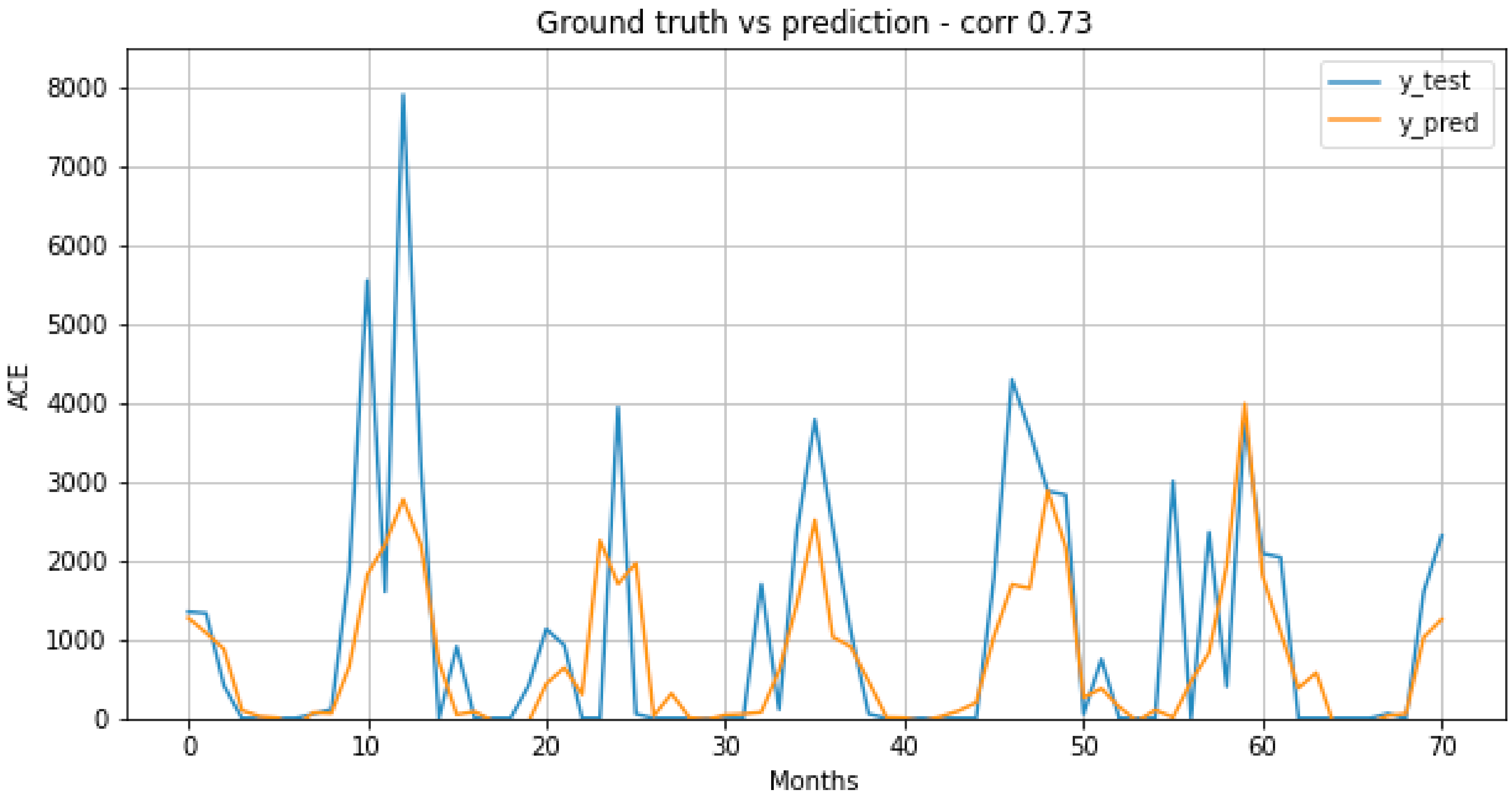


CNN architechture from Fu et al. 2022

Some changes made with respect to original architecture to optimize for the current problem:

- Reduction of the dimension of conv layers
- Added dropout and L2 regularization
- Implementation of early stopping
- Changed loss function to LogCosh
- Change Mean Pooling with Max Pooling

Model 1 (CNN)

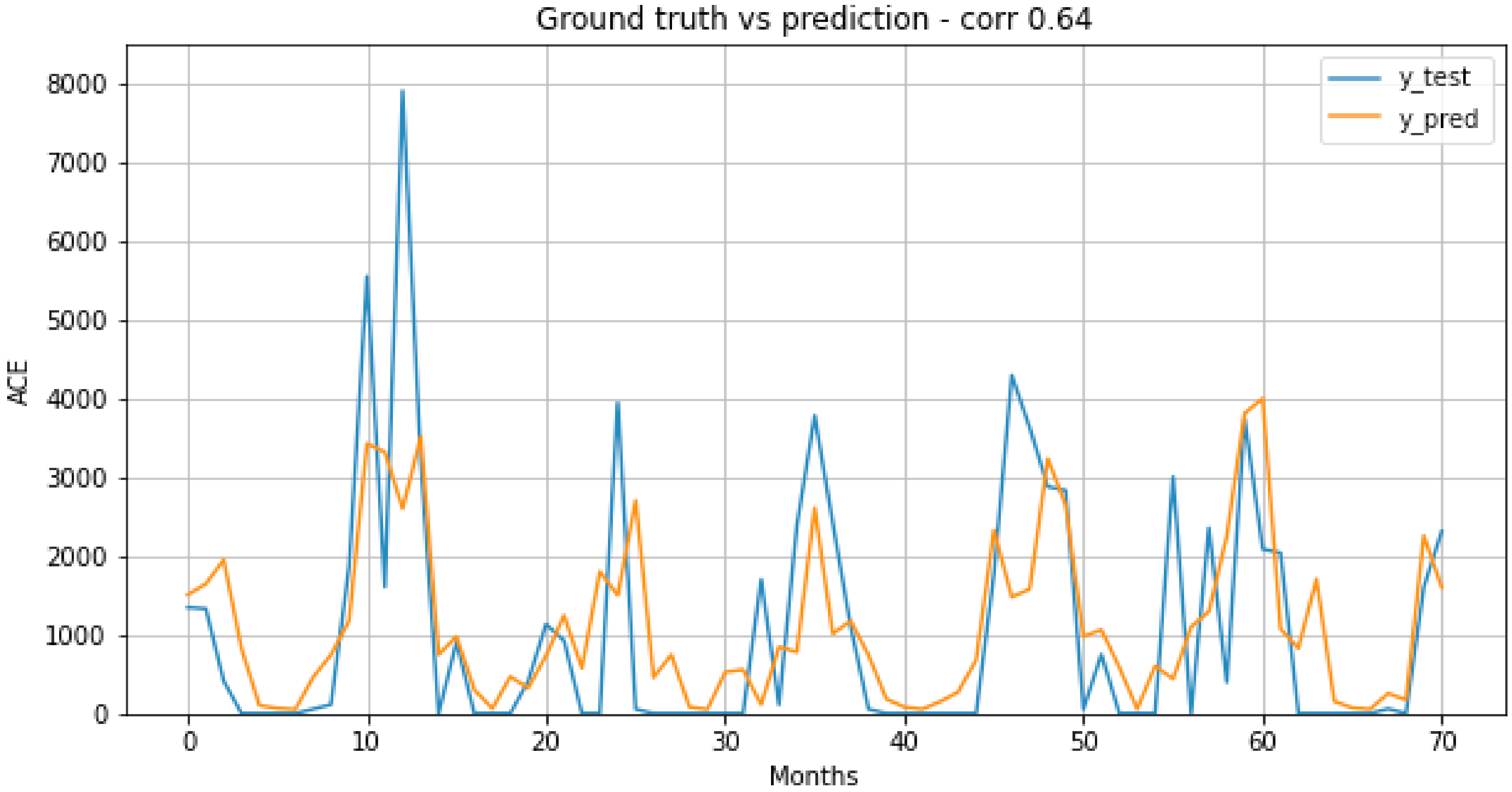


Ground truth vs prediction - corr 0.73
y_test
y_pred
ACE
Months
0
1000
2000
3000
4000
5000
6000
7000
8000
0
10
20
30
40
50
60
70

Model 2 (RF)

Simpler model based on random forest regressor:

- 15 features: spatial averages of the five drivers across western, central and eastern Mediterranean

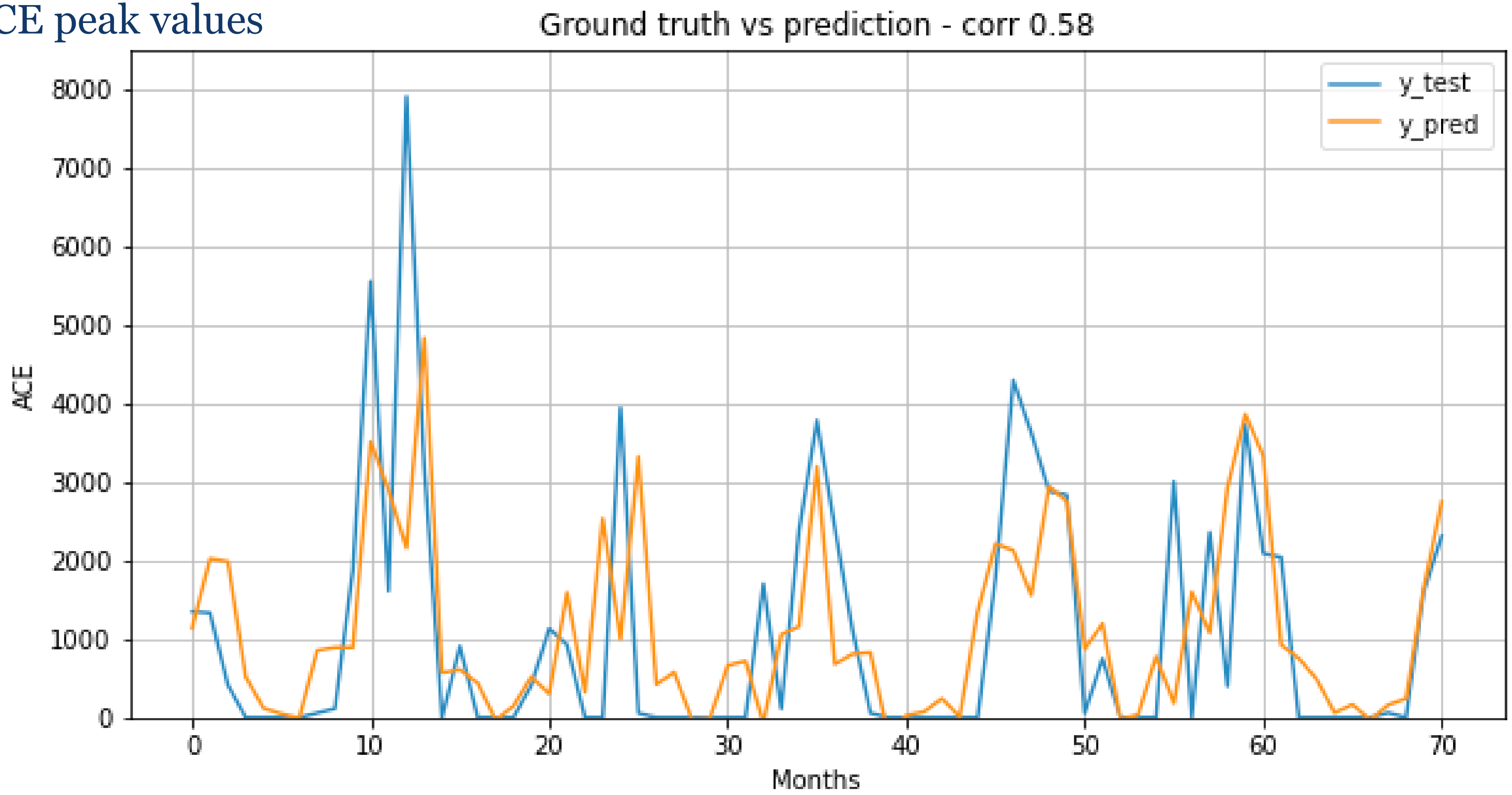


	feat_imp
avo_C	0.441095
rh_E	0.102472
avo_E	0.063274
sst_E	0.051824
avo_W	0.046821
ws_W	0.040147
egr_C	0.040074
ws_E	0.035922
sst_W	0.035149
egr_E	0.033059
egr_W	0.032087
rh_W	0.028478
sst_C	0.026530
ws_C	0.012721
rh_C	0.010347

Model 3 (XGB)

Simpler model based on boosting (XGB algorithm):

- 15 features: spatial averages of the five drivers across western, central and eastern Mediterranean
- Better representation of ACE peak values



Model summary and next steps

CNN	• Good result for correlation • Takes into account spatial patterns	• Not so good skill for the amplitude of ACE peaks
Random Forest	• Easier to interpret	• Needs assumptions on spatial patterns
XGB	• Better representation of peak values	• Lower correlation

- Apply trained models to forecast data
- Best strategy to switch from reanalysis to forecast world? Fine-tuning? Full retraining on hindcast period?

Centro Euro-Mediterraneo sui Cambiamenti Climatici

www.cmcc.it