



Supplement of

Assessing seasonal climate predictability using a deep learning application: NN4CAST

Víctor Galván Fraile et al.

Correspondence to: Víctor Galván Fraile (vgalva01@ucm.es) and Belén Rodríguez-Fonseca (brfonsec@ucm.es)

The copyright of individual parts of the supplement might differ from the article licence.

Supplementary Material

Hyperparameter Dictionary and Preprocessing Code

NN4CAST stores model configurations in a Python dictionary, which can be saved as a YAML file to facilitate reproducibility. The following examples demonstrate the creation of this dictionary, saving it, and preprocessing the data.

```
from tensorflow.keras import activations
import numpy as np
from nn4cast.predefined_classes import Dictionary_saver

5 hyperparameters =
    'path_x' = "/path/to/your/data/HadISST1_sst_1870-2019.nc",
    'path_y' = "/path/to/your/data/HadISST1_sst_1870-2019.nc",

    'time_limits' = [1900,2019],
10    'jump_year' = 0,

    'lat_lims_x' = [30, -30],
    'lon_lims_x' = [120, 290],
    'lat_lims_y' = [40, -10],
15    'lon_lims_y' = [-80, +20],

    'name_x' = 'sst',
    'name_y' = 'sst',

20    'months_x' = [1, 2, 3],
    'months_skip_x' = ['None'],
    'months_y' = [4, 5, 6],
    'months_skip_y' = ['None'],

25    'mean_seasonal_method_x' = True,
    'mean_seasonal_method_y' = True,

    'regrid_degree_x' = 2,
    'regrid_degree_y' = 2,

30    'scale_x' = 1,
    'scale_y' = 1,

    'detrend_x' = True,
35    'detrend_y' = True,
    'detrend_x_window' = 30,
    'detrend_y_window' = 30,
    # Neural network hyperparameters (default ones).
    'layer_sizes' = [1024,256,64,256,1024],
40    'activations' = [activations.elu,activations.elu,activations.elu,
        activations.elu, activations.elu]
    'dropout_rates' = [0.1],
```

```

    'kernel_regularizer' = 'l2',
    'learning_rate' = 0.0001,
    'epochs' = 2500,
    'num_conv_layers' = 0,
57    'use_batch_norm' = True,
    'use_initializer' = True,
    'use_dropout' = True,
    'use_init_skip_connections' = False,
    'use_inter_skip_connections' = False,
62
    'units_x' = ['°C'],
    'units_y' = ['°C'],
    'region_predictor' = 'Tropical Pacific',
    'p_value' = 0.1,
67    'outputs_path' = "/path/to/the/directory/Outputs_ND_sst_SO/"

Dictionary_saver(hyperparameters)

```

Listing S1. Python code example demonstrating how to create an instance of the parameter and hyperparameter dictionary, and how to save it using the `Dictionary-saver` function.

- 5 The preprocessing of predictor and predictand data is carried out with the `Preprocess` function, which executes the operations defined in the hyperparameter dictionary, including regridding, computing seasonal anomalies, and detrending:

```

from nn4cast.predefined_classes import Preprocess

dictionary_preprocess = Preprocess(dictionary_hyperparams=hyperparameters)

```

Listing S2. Python code example demonstrating the use of the `Preprocess` function to manipulate predictor and predictand data, and to store the processed outputs in a dictionary format.

- Detrending is implemented with a backward moving average (BMA) algorithm, where the mean over a sliding window of previous years is subtracted from each subsequent value. The window length is specified in the hyperparameters as `detrend_x_window` and `detrend_y_window`. This method ensures that only past information is used and avoids
- 10 introducing future knowledge into the preprocessing phase.

Model Cross-Validation and Explainable AI

- NN4CAST implements a cross-validation procedure using the `Model_build_and_test()` function. The number of folds is defined by `n_cv_folds`, and a fraction of each training fold (10%) is reserved for early stopping. This procedure ensures robust estimates of model performance and can be configured for leave-one-out cross-validation if desired. Attributions for a
- 15 target region are computed using the Integrated Gradients methodology from the Alibi library (Klaise et al., 2021), leveraging

the cross-validation results to cover the entire hindcast period. To compute these attributions for a specific target region of the predictand field, first set the instance to `True` to enable the computation. Then, define the region as a list of latitude and longitude ranges, using the parameter (setting `region_importances=[[latitude_min, latitude_max], [longitude_min, longitude_max]]`).

```
from nn4cast.predefined_classes import Model_build_and_test, Results_plotter

outputs_cross_validation = Model_build_and_test(dictionary_hyperparams =
    hyperparameters, dictionary_preprocess=dictionary_preprocess,
5    cross_validation=True, n_cv_folds=120, plot_differences=False, importances=True,
    region_importances=[[5,25], [-55,-15]])

Results_plotter(hyperparameters, dictionary_preprocess, rang_x=1.5, rang_y=1,
10    predictions=outputs_cross_validation['predictions'],
    observations=outputs_cross_validation['observations'],
    years_to_plot=[2015], plot_with_contours=True,
    importances=outputs_cross_validation['importances'],
    region_importances=outputs_cross_validation['region_attributed'])
```

Listing S3. Python code example demonstrating the use of the `Model-build-and-test` function to construct the model using selected hyperparameters, followed by training, validation, and testing through a user-defined cross-validation approach. The listing also includes the use of the `Results-plotter` function to visualize the model outputs and their attributions for the target region over the predictor field..

Hyperparameter Optimization

- 20 Hyperparameter tuning is carried out with the `Model_searcher()` function, which explores a user-defined search space using a random search methodology (Bergstra and Bengio, 2012). The best combination is selected based on cross-validation performance.

```

from nn4cast.predefined_classes import Model_searcher

params_selection =
    'pos_number_layers' = 5, # set the maximum value of fully connected layers (int).
5    'pos_layer_sizes' = [16, 64, 256], # set the possible layer sizes (list).
    'pos_activations' = ["elu", "linear"], # set the possible activation functions
        (possibilities are all the ones available: tf.keras.layers.activations()) (list).
    'pos_dropout' = [0.0, 0.01], # set the possible dropout probabilities (list).
10    'pos_kernel_regularizer' = ["l1_l2"], # set the possible kernel regularizer
        (possibilities are: l1_l2, l1, l2, None) (list).
    'search_skip_connections' = False, # set if searching for skip connections, either
        intermediate or end_to_end connections (bool).
    'pos_conv_layers' = 0, # set the maximum number of convolutional layers, the
        predictor field (X) must be 2D (int).
15    'pos_learning_rate' = [1e-4, 1e-3], # set the possible learning rates (list).

outputs_bm_cross_validation = Model_searcher(dictionary_hyperparams=hyperparameters,
        dictionary_preprocess=dictionary_preprocess,
        dictionary_possibilities=params_selection, max_trials=10, n_cv_folds=8)

```

Listing S4. Python code example demonstrating the use of the `Model-searcher` function to identify the optimal set of hyperparameters within a predefined search space. The hyperparameter ranges are based on commonly reported values in the specialized literature and explored using the random search method (Bergstra and Bengio, 2012). The best combination obtained is then evaluated following the selected cross-validation scheme.

Application 1: Modelling Pacific-North Tropical Atlantic SST Connection

The `Results-plotter` function enables visualization of the model predictions for specific cases, allowing for direct comparison with the observed data. Two representative years—1986 and 1998—have been selected for this purpose and are illustrated in Fig. S1 and Fig. S2, respectively.

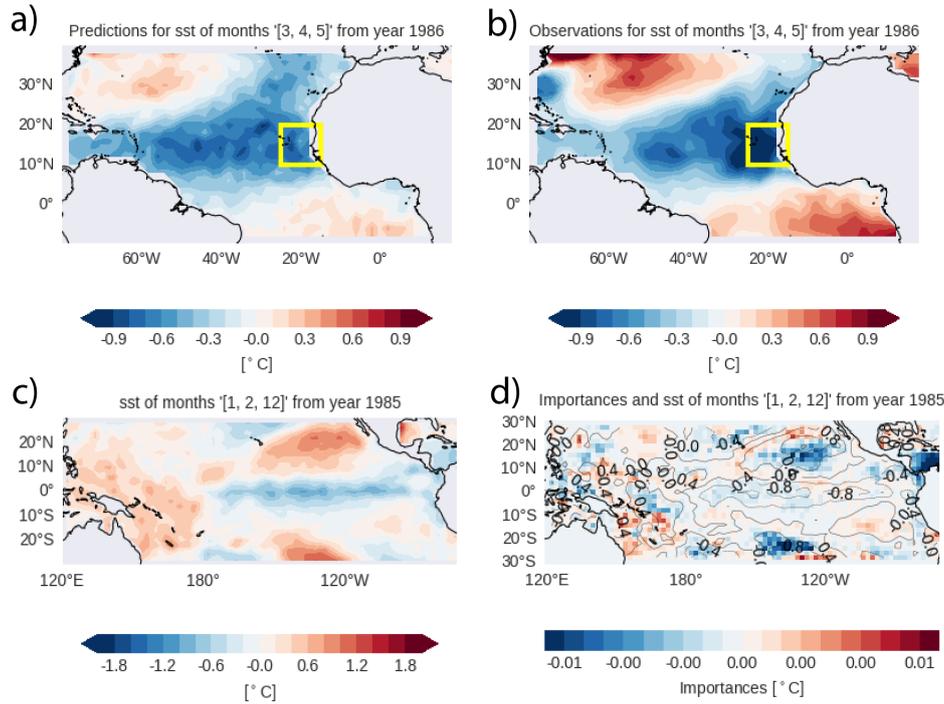


Figure S1. Comparison of model predictions for a specific sample year. Panels show: (a) predicted MAM Atlantic SST anomalies ($^{\circ}\text{C}$); (b) observed MAM Atlantic SST anomalies ($^{\circ}\text{C}$) for the year 1986; (c) Pacific SST during DJF ($^{\circ}\text{C}$) for the year 1985 used as the predictor field; and (d) Pacific SST DJF predictor field for the 1985 sample, with contours ($^{\circ}\text{C}$) representing SST values and shaded areas ($^{\circ}\text{C}$) indicating the model attributions for the selected predictand region (yellow box in panels a and b). The yellow box marks the target region where model attributions are computed.

The first case presented (Fig. S1) corresponds to a central La Niña event, characterized by negative SST anomalies in the central equatorial Pacific, along with positive anomalies over the Warm Pool and subtropical regions of the Pacific Ocean. In this scenario, the model predicted a cooling of the WTNA and the SMSCU region, as well as a warming over the eastern equatorial Atlantic and the Gulf Stream region. To gain insight into how the model made these predictions, we focused on the SMSCU region (highlighted by a yellow box in Fig. S1a) and Fig. S1b) and computed the Integrated Gradients attributions for the 1986 prediction. These attributions are shown in Fig. S1d). The importance values are expressed in units of the predicted variable ($^{\circ}\text{C}$), which in this case matches the units of the predictor (SST anomalies). The attribution map reveals the individual contribution of each SST predictor to the prediction over the region marked by the yellow box. Notably, both the Caribbean region and the warming zones in the central tropical Pacific greatly contributed to the prediction of negative SST anomalies in the Mauritanian-Senegalese Coastal Upwelling region.

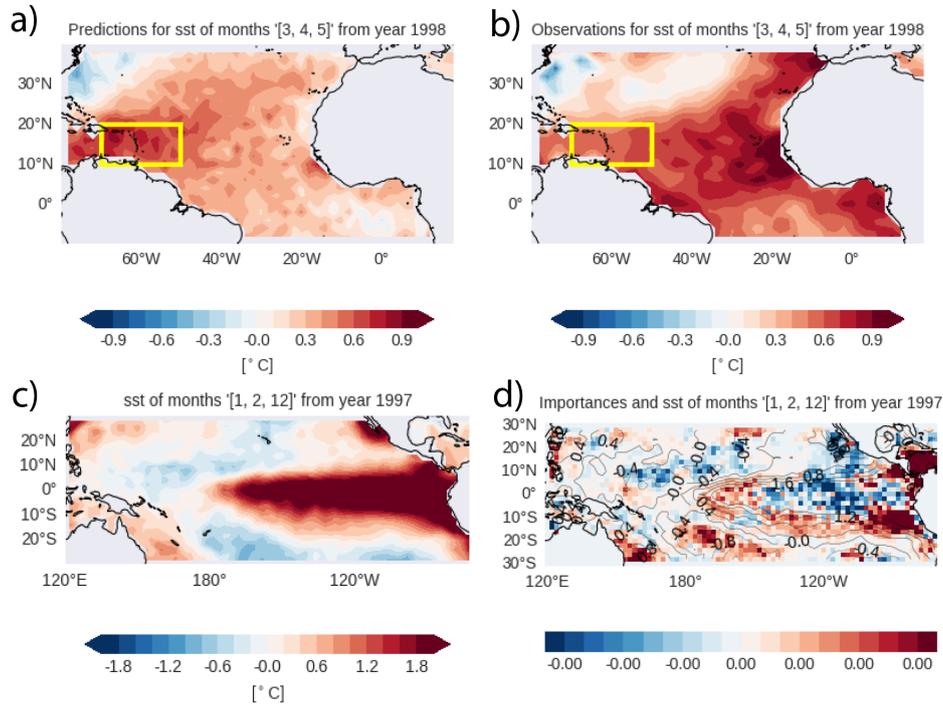


Figure S2. Comparison of model predictions for a specific sample year. Panels show: (a) predicted MAM Atlantic SST anomalies ($^{\circ}\text{C}$); (b) observed MAM Atlantic SST anomalies ($^{\circ}\text{C}$) for the year 1998; (c) Pacific SST during DJF ($^{\circ}\text{C}$) for the year 1997 used as the predictor field; and (d) Pacific SST DJF predictor field for the 1997 sample, with contours ($^{\circ}\text{C}$) representing SST values and shaded areas ($^{\circ}\text{C}$) indicating the model attributions for the selected predictand region (yellow box in panels a and b). The yellow box marks the target region where model attributions are computed.

Figure S2 illustrates the second case, corresponding to a canonical El Niño event in 1998, associated with widespread warming in the tropical Atlantic, including the WTNA region. In this case, the model accurately reproduced the observed warming signal. For the attribution analysis, we selected a classic WTNA index region, delineated by the yellow box in Fig. S2a) and Fig. S2b).

40 The attribution map for this case is somewhat noisier compared to the 1986 scenario. Nevertheless, some consistent patterns emerge: there are positive contributions from the Caribbean, the central Pacific, and a latitudinal band south of 10°S . In contrast, negative contributions are found mainly in the Niño 3 region and certain areas north of 10°N , which act to counterbalance the warming signal.

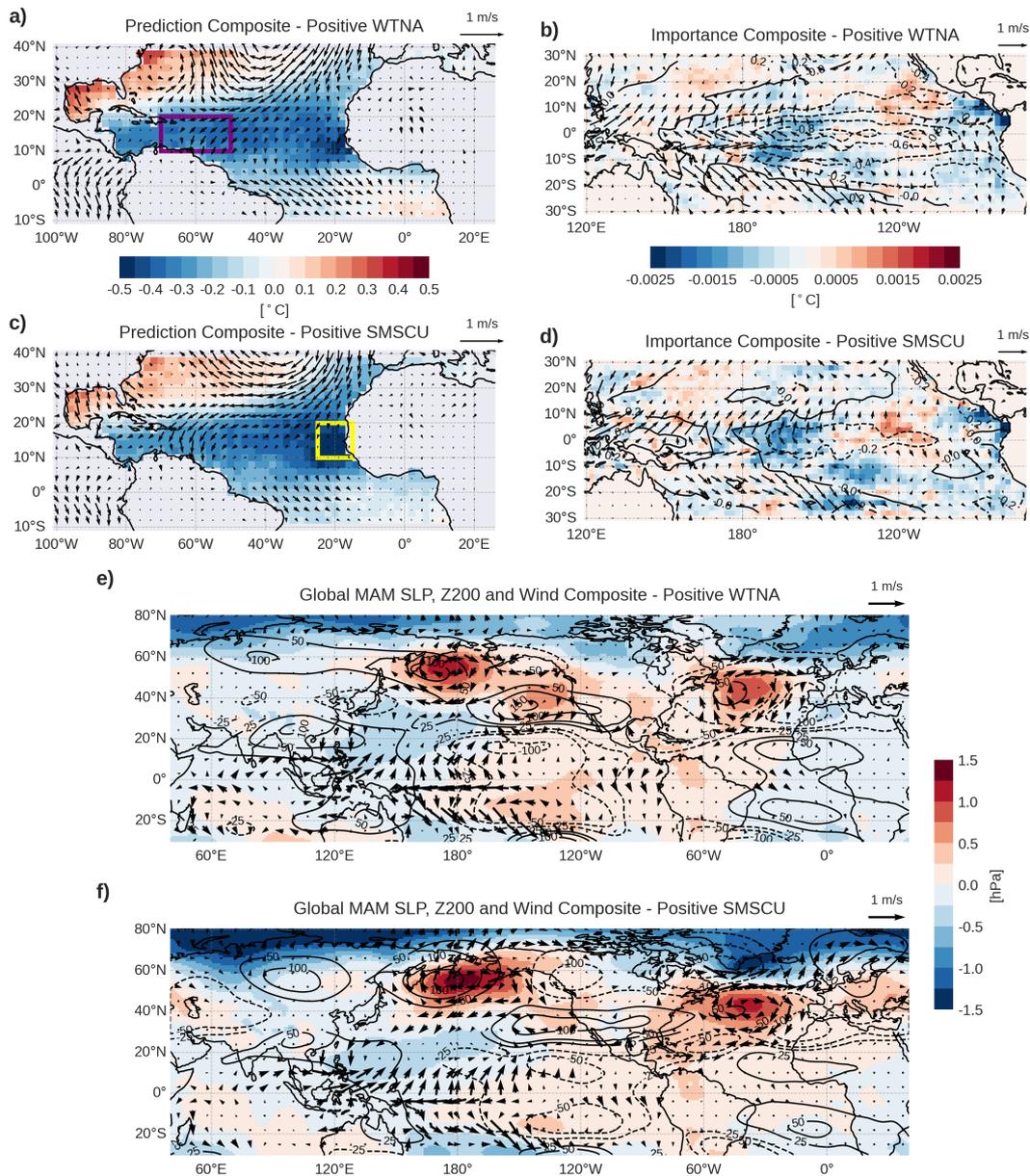


Figure S3. Composites of model anomalous SST predictions, predictor fields, and attribution maps for negative predicted WTNA and SMSCU, based on 28 and 29 events, respectively. Panels (a) and (c) show the predicted mean SST anomalies in the Atlantic during MAM together with surface wind anomalies indicated by arrows. Panels (b) and (d) show the attribution maps over the predictor fields with SST in contours and surface winds in arrows. Panels (e) and (f) display global composites of MAM anomalies in sea level pressure (shading), 200 hPa geopotential (contours), and surface winds for positive WTNA and SMSCU events. Attribution maps indicate the relative contribution of each grid point in the predictor field to the forecasted value in the target region, with the sum of the values within each map matching the predicted anomaly in the corresponding index region (i.e., the sum of values in panel (c) matches the WTNA anomaly within the purple box in panel (a)).

Figure S3 represents the same as Figure 4 but for the negative events. Composites based on the Niño 3.4 index are also
45 presented in Figure S4 A traditional approach of teleconnection analysis is the regression of the indices over the SST fields. This
is illustrated for comparison in Figure S5.

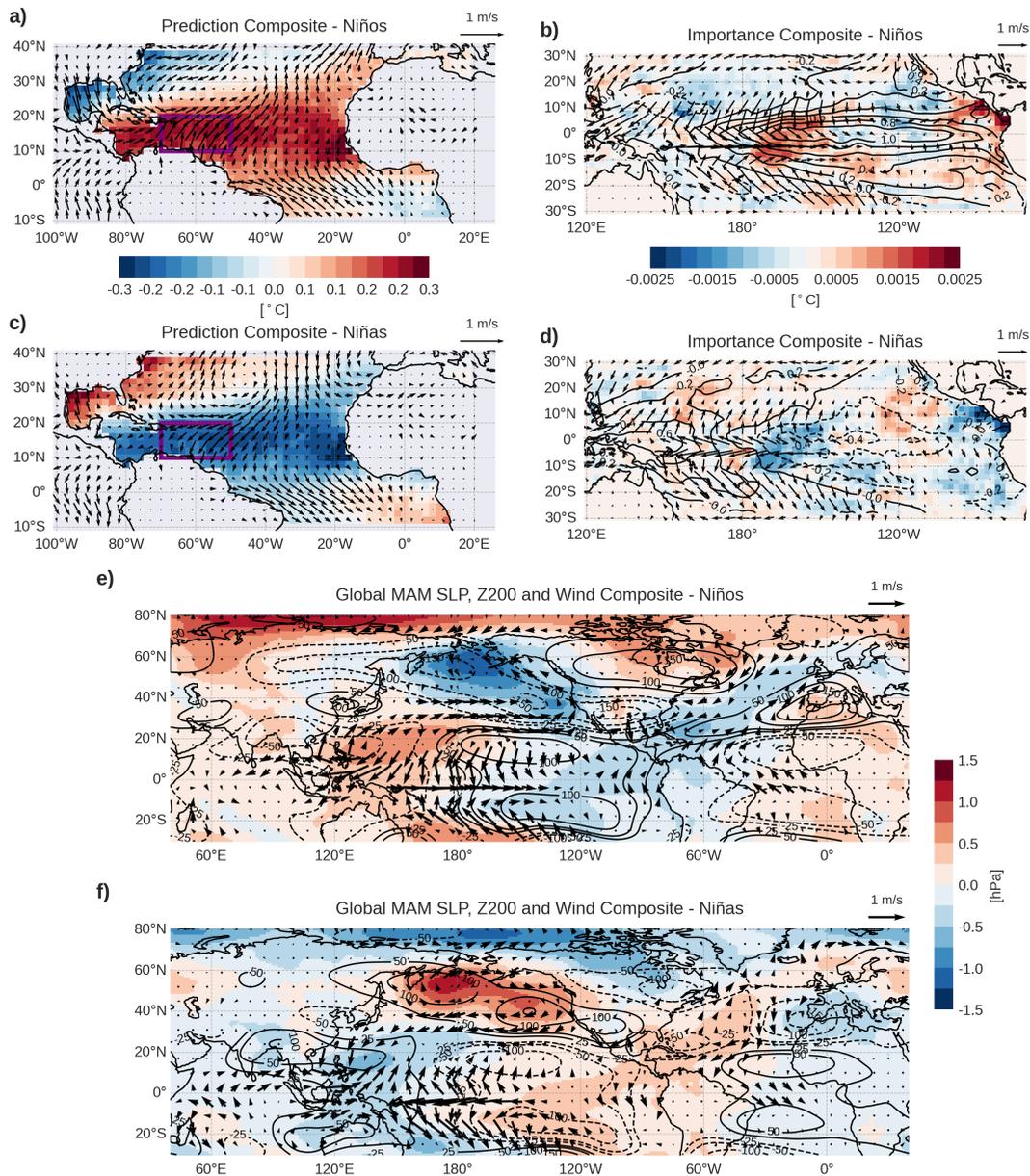


Figure S4. Composites of model anomalous SST predictions, predictor fields, and attribution maps for Niño 3.4 positive (Niño) and negative (Niña), based on 35 and 34 events respectively. Panels (a) and (c) show the predicted mean SST anomalies in the Atlantic during MAM together with surface wind anomalies indicated by arrows. Panels (b) and (d) show the attribution maps over the predictor fields with SST in contours and surface winds in arrows. Panels (e) and (f) display global composites of MAM anomalies in sea level pressure (shading), 200 hPa geopotential (contours), and surface winds for Niños and Niñas. Attribution maps indicate the relative contribution of each grid point in the predictor field to the forecasted value in the target region, with the sum of the values within each map matching the predicted anomaly in the corresponding index region (i.e., the sum of values in panel c matches the WTNA anomaly within the purple box in panel (a)).

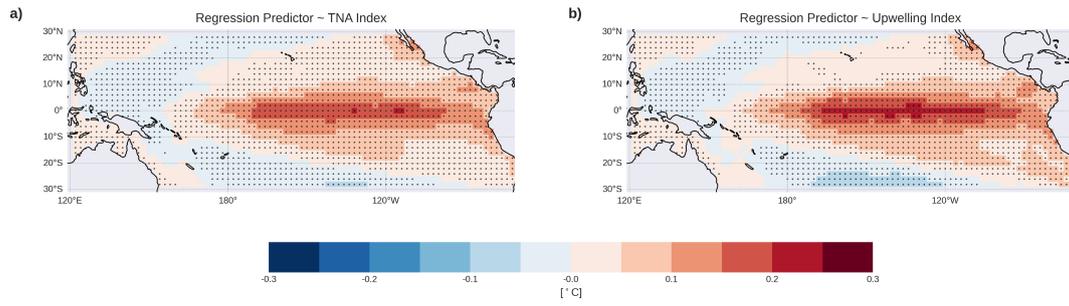


Figure S5. Regressions of WNTA and SMSCU indices in MAM over the DJF SST anomalies. Statistically significant results, based on a two-tailed t-test with a significance level defined in the hyperparameters (95%), are indicated by the dotted regions.

Application 2: Modelling North and Tropical Pacific - European Precipitation Teleconnection

Figure S6 shows the same information as Figure 7, but for the composite of predicted negative precipitation anomalies.

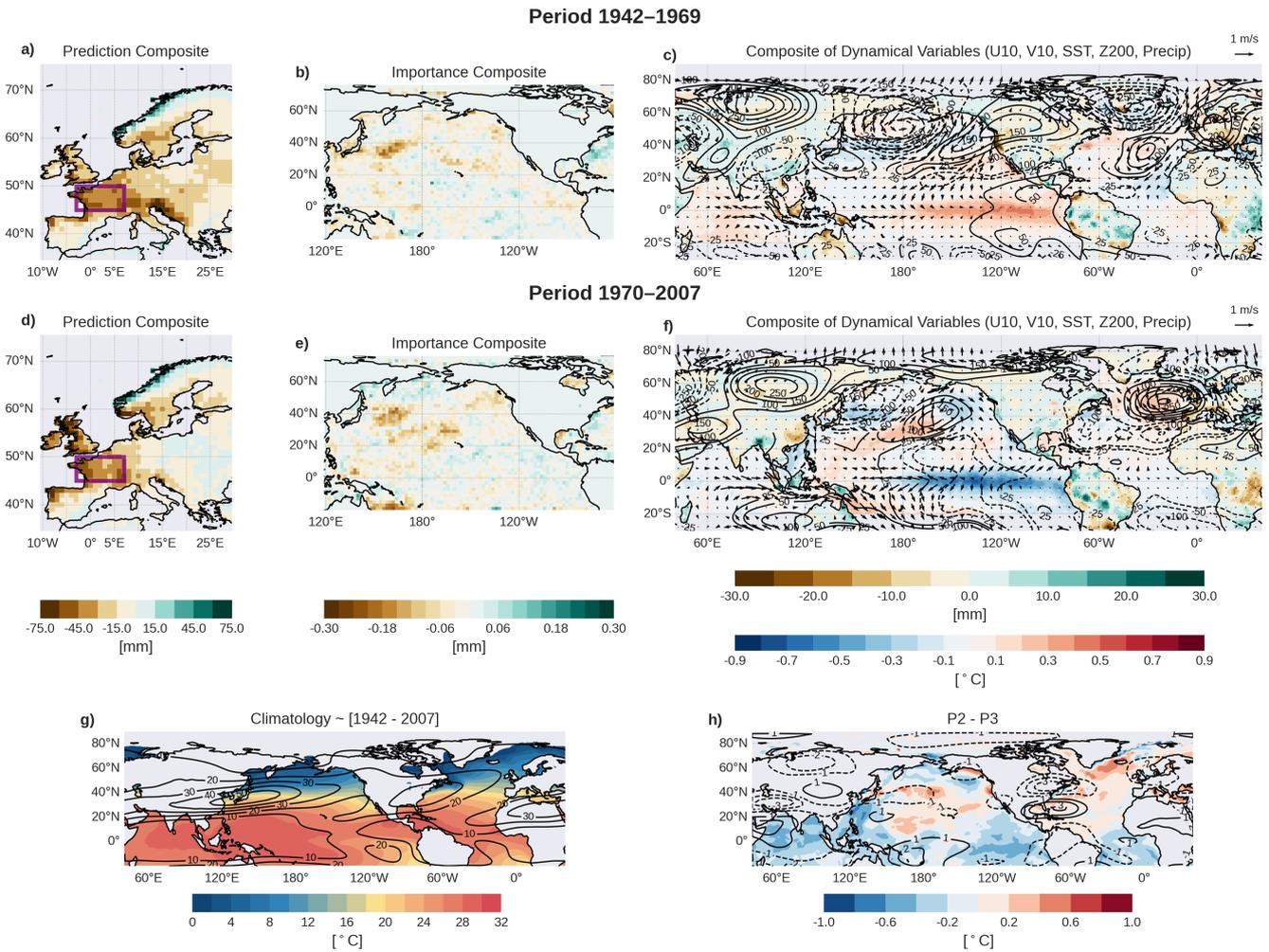


Figure S6. Composites of model predictions, predictor fields, and attribution maps based on negative precipitation events in western central Europe (index defined as the purple rectangle in (a)), with 11 events in period P2 [1942-1969] and 11 events in period P3 [1970-2007]. Panels show: (a, d) precipitation anomalies in Europe during OND for periods P2 and P3, respectively; (b, e) attribution maps over the predictor field corresponding to negative events for periods P2 and P3, respectively. Panels (c, f) display global composites of OND anomalies in sea surface temperature and precipitation (shading), 200 hPa geopotential (contours), and surface winds for positive events for periods P2 and P3, respectively. Attribution maps (b, e) indicate the relative contribution of each grid point in the predictor field to the forecasted value in the target region. The sum of the attribution values within each map equals the predicted anomaly in the corresponding index region (i.e., the sum of values in panel (b) matches the precipitation anomaly within the purple box in panel (a)). Panels (g, h) show the climatology of the SST and U200 for the period [1942-2007] as well as the differences from periods P2 and P3.