SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Innovative iNitialisation techniques for multi-annual ClImate PredIcTions (INCIPIT)
Computer Project Account:	spitvolp
Start Year - End Year :	2019 - 2020
Principal Investigator(s)	Danila Volpi
Affiliation/Address:	Institute of Atmospheric Sciences and Climate, National Research Council (ISAC-CNR) Via Piero Gobetti 101, 40129 Bologna, Italy
Other Researchers (Name/Affiliation):	ISAC-CNR: Susanna Corti BSC: Francisco Doblas-Reyes

The following should cover the entire project duration.

Summary of project objectives

(10 lines max)

The objective of this project is to test specific improvements to the standard anomaly initialisation technique to assess whether addressing some of the limitations detected in the technique has a positive impact on the skill of multi-annual predictions. The limitations to tackle are namely:

- the risk of introducing anomalies recorded in the observed data whose amplitude does not fit in the range of the internal variability generated by the model. This could result in the model erroneously predicting extreme events, such as an intense El Niño or a pause in the thermohaline circulation (first year experiment);
- the displacement between the geographical position where the model develops the variability modes,and the actual observed position of the modes, which could lead to a wrong propagation of the information coming from the observed state (second year experiment).

Summary of problems encountered

(If you encountered any problems of a more technical nature, please describe them here.)

Since the preliminary results of the quantile matching experiment were showing promising findings especially avoiding the deep convection collapse in the Labrador Sea which characterises the EC-Earth DCPP experiment initialised with full field, we have decided to use the resources available the second year to enlarge the ensemble size of the quantile matching experiment in order to achieve a more robust comparison. The quantile matching decadal experiment is composed by a 10-member ensemble with start dates from 1960-2014.

Experience with the Special Project framework

(Please let us know about your experience with administrative aspects like the application procedure, progress reporting etc.)

The progress reports requirements and the application procedure has been very clear and smooth. Very good assistance by HPC staff has also been provided.

Summary of results

(This section should comprise up to 10 pages, reflecting the complexity and duration of the project, and can be replaced by a short summary plus an existing scientific report on the project.)

The quantile matching initialisation method aims at tackling two major issues:

• The drift coming from initialising the predictions away from the model preferred trajectories; this is in also the objective of any other anomaly initialisation method that employs an initial state that belongs to the model attractor and only imposes the observed variability (Smith et al. 2008; Pohlmann et al. 2013);

• The potential inconsistencies between the observed/model distribution of variability.

The QM method selects the initial condition of the prediction as the model state which is identically located in the model distribution as the observed initial state in the observed distribution. Therefore, the initial state belongs to the model attractor, reducing the drift. Moreover, matching the observed and model statistical distributions scales the observed variability toward the model one, correcting any potential amplitude incompatibilities.

The model in use for this study is the CMIP6 version of EC-Earth3 GCM (Döscher et al. 2021). The benchmark hindcasts are a full field initialized experiment (FFI) and an ensemble of 15 uninitialised CMIP6 historical simulations (Histo), also assessed by Bilbao et al. (2021).

Both initialised experiments (FFI and QM) are composed by 10 ensemble members, initialised every November from 1960 until 2014 and running for 5 years.

The ensemble is generated using perturbations in the 3-dimensional temperature field in the atmospheric component, and using the 5-member ensemble from the NEMOVAR-ORAS4 reanalysis in the ocean.

The QM is applied to the ocean component and is performed for all the ocean prognostic variables and grid points, by matching the 5 members of NEMOVAR-ORAS4 separately, in order to obtain 5 different initial conditions. The atmospheric and sea-ice components of the QM experiment have identical initial condition as the FFI.

The main scientific question is to evaluate the effect of the drift reduction on the forecast skill. The prediction skill assessment has led to the following findings:

• The skill in surface temperature in the first forecast year is globally significant, the main exceptions being a few regions over Asia and the Southern Ocean and a small region in the North Atlantic. At longer forecast times (2-5 forecast years), the region of non-significant skill in the North Atlantic expands, as well as in the Southern Ocean, and a new region in the central Tropical Pacific loses significance.

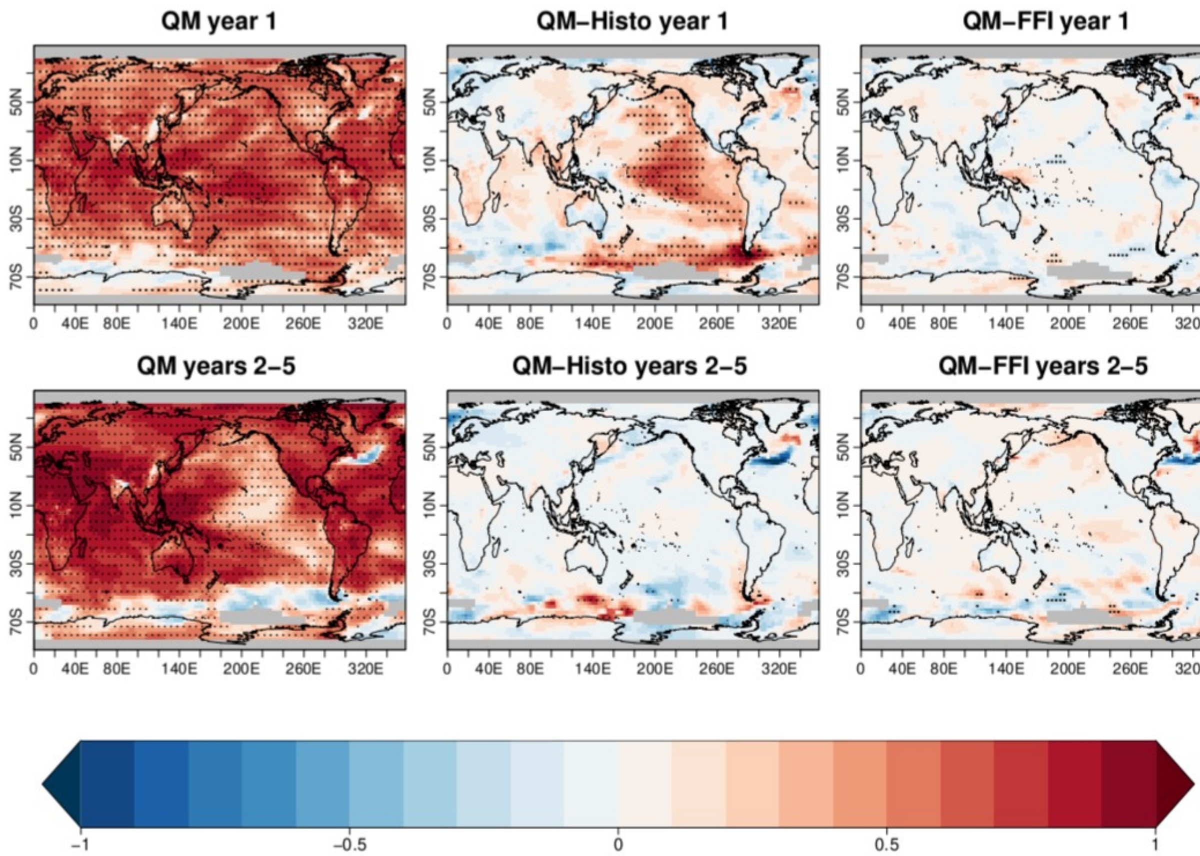

Figure 1. Anomaly correlation for the surface temperature calculated against the GISTEMPv4 dataset. The first column represents the quantile matching (QM) skill in the first forecast year (top) and the 2-5 forecast years (bottom). The middle column shows the difference between QM and Histo skill, and the third column the difference between QM and full-field initialization (FFI) skill. The black dots indicate regions of significant anomaly correlation (AC) (first column), and regions where the difference in AC is statistically significant with 95% confident level (second and third column).

When compared to Histo, the QM predictions show significant improvements over a predominant part of the Pacific and over the North Atlantic subpolar region (NASP) as well as the Southern Ocean, during the first forecast year (top-central panel Figure 1). The significant added-value with respect to Histo in the NASP surface temperature is maintained at longer forecast years (bottom-central panel of Figure 1). However, the improvement is partially lost at forecast years 2-5 in the zone of the inter-gyre position, where the subpolar and subtropical gyres meet. The QM also outperforms the FFI in predicting the NASP surface temperature and, more importantly, the improved skill maintains significance throughout the whole forecast time (third column Figure 1). However, analogously to what is shown in the comparison with Histo, there is a degradation of skill along the Gulf stream for the forecast years 2-5, which could potentially be due to a wrong positioning of the current in the QM predictions.

• Figure 2 shows the skill comparison between QM and Histo/FFI for the upper layer ocean heat content (0–300 m). Similarly to what is shown for the surface temperature, the QM significantly improves the skill in the Pacific and the NASP region with respect to Histo, although there is not a clear improvement over the Southern Ocean. Besides, skill over the Indian Ocean is improved. Part of those improvements are also maintained at longer timescales. When comparing with the FFI prediction, we find that the added value of the QM is mainly located over the NASP region, consistently with the surface temperature results.

The third panel of Figure 3 zooms over the western subpolar North Atlantic region (50–65° N, 60–30° W, WSPNA), as this is a region in which SST and ocean heat content have been shown to influence the temperature and precipitation over the neighbouring continents. The highest values of anomaly correlations -ACs- and the lowest values of root mean square error (RMSE) over the WSPNA region and during the entire forecast time are obtained with the QM (red lines). Conversely, the Histo simulations (yellow lines) do not show significant AC in any of the forecast years, whereas the FFI loses its skill after the first forecast year.

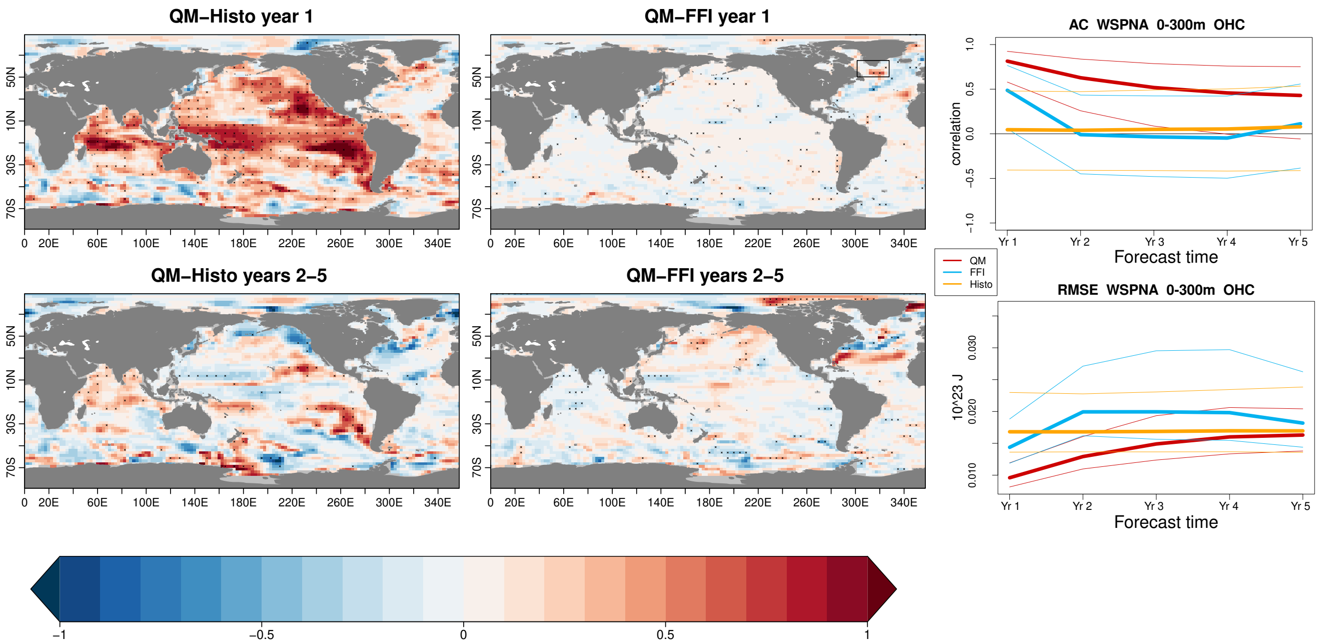

Figure 2. Skill difference in the upper 300 m ocean heat content, computed against EN4 observational dataset. The first column shows the anomaly correlation (AC) difference between quantile matching (QM) and Histo for the first forecast year **(top)** and the 2–5 forecast years **(bottom)**. The second column shows the difference between QM and full-field initialization (FFI). The third column is a zoom on the regional mean of the western subpolar North Atlantic sector (WSPNA; 50–65° N, 60–30° W) as indicated by the black box. The AC and root mean square error (RMSE) are calculated with yearly mean data along the forecast time and are shown, respectively, in the **top** and **bottom** panel. In red is shown the QM experiment, in blue the FFI, and in yellow the Histo. The thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a χ^2 distribution for the RMSE.

• One of the most remarkable result is that while the full field decadal predictions experience a collapse of deep convection in the Labrador Sea (Bilbao et al. 2021), the quantile matching predictions manage to avoid this problem as shown in Figure 3. Therefore, this could be due to the positive effect of initialising the predictions on the model attractor. However, this does not directly translate into an increase of forecast skill, as the QM does not improve with respect to the Histo ensemble that preserves the convection. While Histo has a constant skill throughout the whole forecast time, QM presents statistically significant AC only during the first year (top left panel of Figure 4). This might be due to the fact that the skill in predicting the mixed layer depth is dominated by the external forcing changes that have induced an increase in stratification hindering convection activity. The additional model variability that attempts at capturing the observed one rather constitutes an additional noise, which is detrimental to the forecast performance.

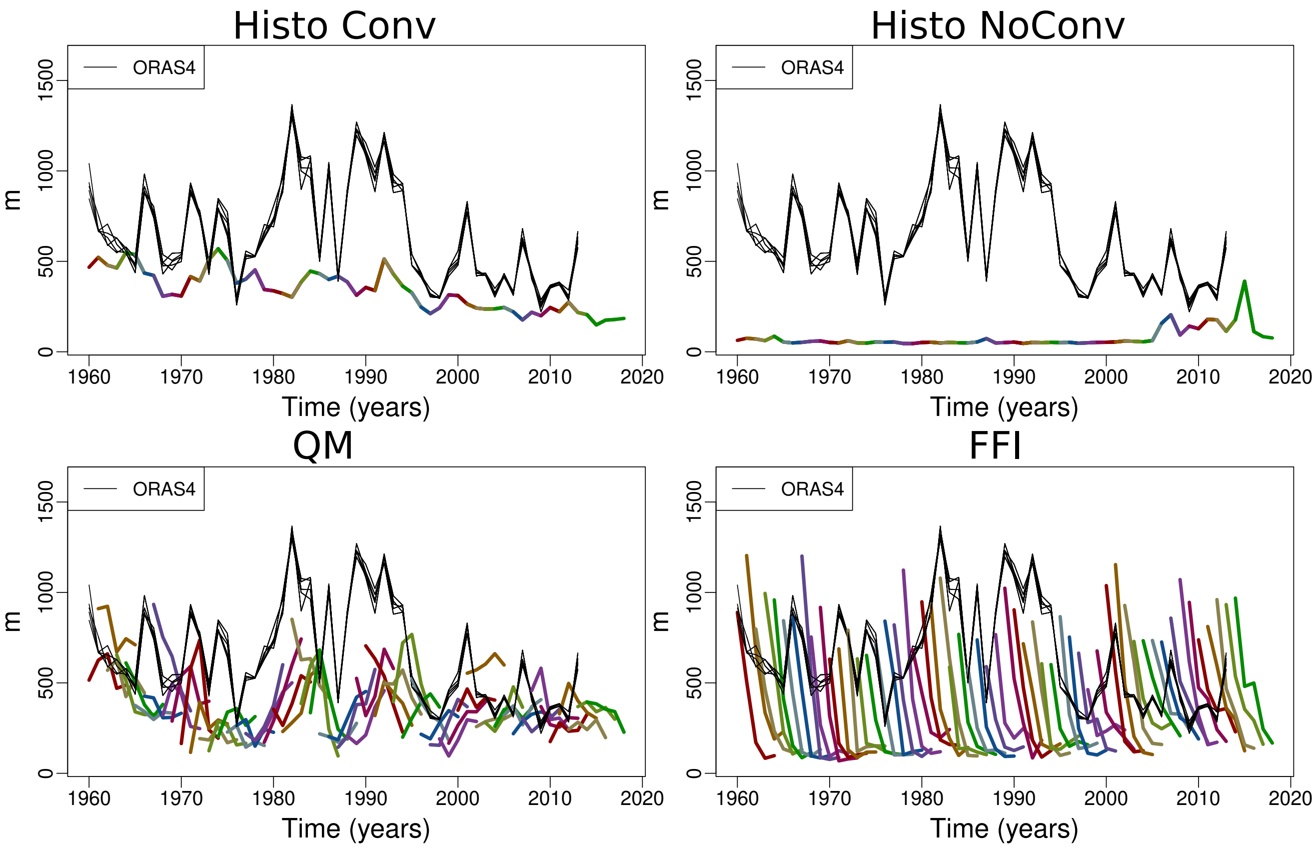

Figure 3. Evolution of the mixed layer depth ensemble mean, as a proxy for the deep convection intensity for the February-March-April average in the Labrador Sea. The top row shows the Histo ensemble splitted into members that exhibit a regular convection activity **(left)**, from the three which show no convection throughout most of the historical period **(right)**. The second row shows, respectively, the QM **(left)** and the FFI **(right)** evolution. The different start dates are represented with different colors. The NEMOVAR-ORAS4 is shown in black.

• To assess the skill in predicting the barotropic circulation we look at the barotropic stream function, which characterises the horizontal water transport integrated vertically in the Labrador Sea (right column of Figure 4). We find that Histo do not show any skill. The initialised predictions, on the other hand, have similar skill at the beginning of the forecast. However, the skill of FFI deteriorates with forecast time and is lost at forecast year 3, while the skill of QM is maintained throughout the whole forecast time.

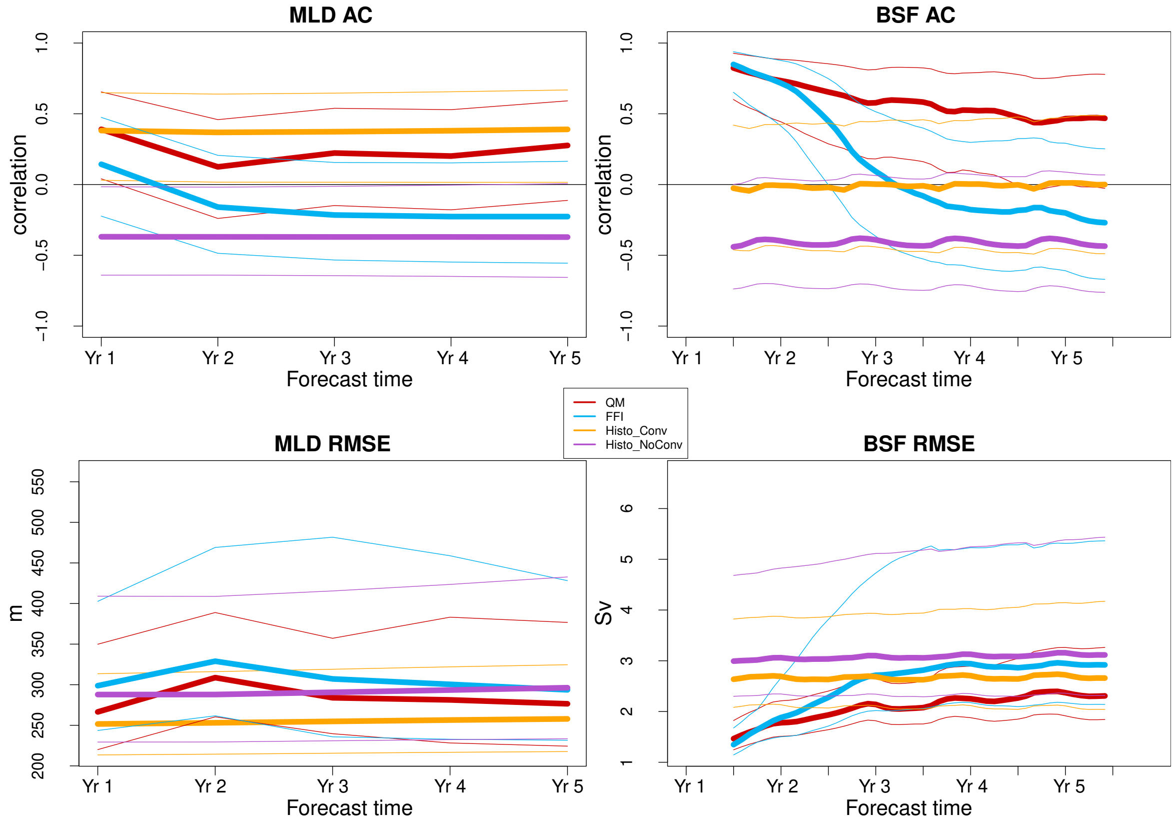

Figure 4. Anomaly correlation of the mixed layer depth (**top left** panel) and the barotropic stream function (**top right** panel) in the Labrador Sea (55–65°N, 56–46°W). The second row shows the respective root mean square error (RMSE). Quantile matching (QM) is shown in red, the full-field initialization (FFI) in blue, the Histo with convection is shown in yellow, and in purple the Histo members with no convection. The skill are computed against NEMOVAR-ORAS4, using the February–March–April mean for the mixed layer depth, and a running mean of 12 months for the barotropic stream function. The thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a χ^2 distribution for the RMSE.

• The correct representation of the deep convection in the Labrador Sea also has an impact on the Atlantic Meridional Overturning Circulation, which is skillfully predicted by the QM throughout the whole forecast time as shown in the Hovmoller diagrams of Figure 5. If we focus north of 40°N, the FFI starts with high skill that progressively deteriorates until the third forecast year down to an AC smaller than 0.1, probably due to the collapse in deep convection. This is largely improved by the QM that shows high skill in predicting the AMOC, particularly at subpolar latitudes, for the entire forecast time. The skill of FFI at lower latitudes is marginally higher than the QM one throughout the whole forecast time. The skill of the Histo simulations is constant with forecast time. This is consistent with the results in Borchert et al. (2021), where the good representation of the NASP SST in the CMIP6 historical simulations is attributed to the AMOC-related response to the forcings (volcanic eruptions and partly solar forcing).

• We complement the North Atlantic analysis with the assessment of the skill in predicting the AMV. The ensemble QM predictions successfully capture the observed warming trend (left panel of Figure 6). In terms of skill, both the initialised predictions and the Histo simulations succeed in predicting the AMV index with skill significantly different than 0 along the whole forecast period as shown by the confidence intervals (thin lines in Figure 6). The prediction skill of FFI starts close to the skill of Histo and it might be generated by the AMV persistence. During the first two forecast years, QM outperforms FFI, both in terms of AC and RMS (second and third panel of Figure 6), although the difference is not statistically significant.

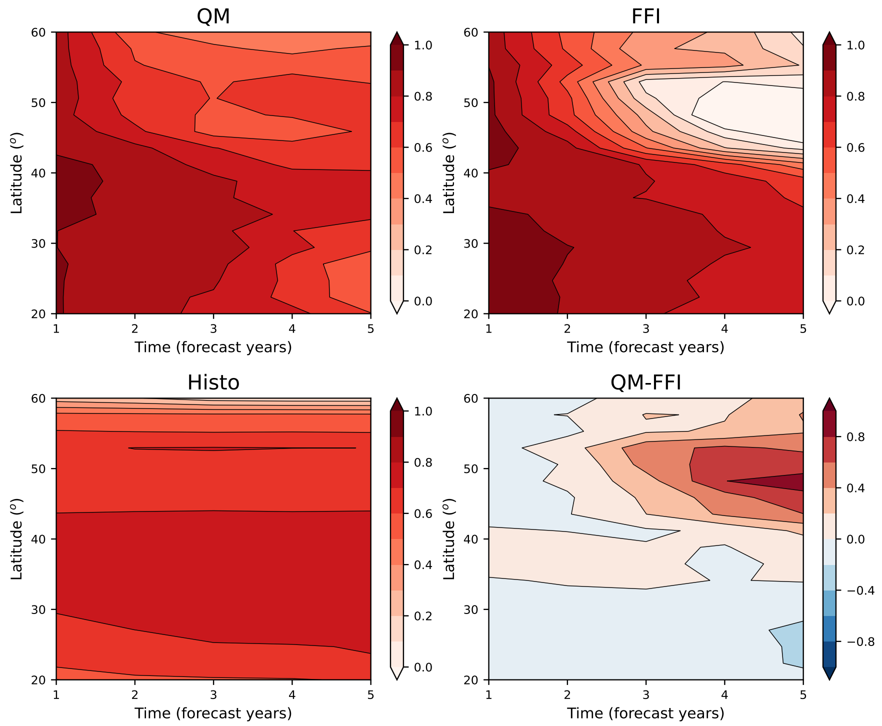

Figure 5. Hovmoller diagrams of the AMOC anomaly correlation with NEMOVAR-ORAS4. The Atlantic Meridional Overturning Circulation (AMOC) is calculated as the maximum of the Atlantic meridional overturning stream function over adjacent boxes of 2° latitude (from 20° to 60° N) and 900-3000 m depth. The panels show, respectively, the quantile matching (QM) **(top left)**, the full-field initialization (FFI) **(top right)**, the Histo **(bottom left)**, and the difference between the QM and the FFI skill **(bottom right)**.

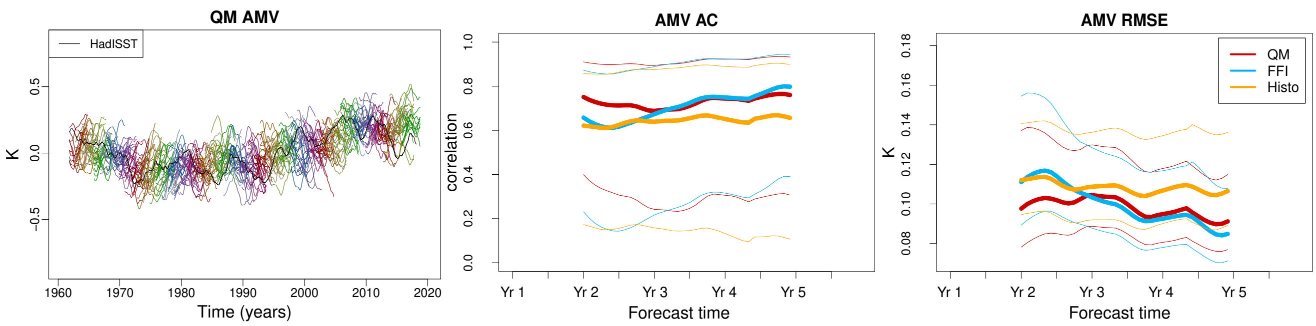

Figure 6. Atlantic multidecadal variability (AMV) index for the quantile matching (QM) predictions **(left)**. The index computed from HadISST is shown in black. The different colors represent the different start dates. The anomaly correlation (AC) and the root mean square error (RMSE) of the AMV index are shown, respectively, in the **central** and **right** panels. The skill scores are calculated against the HadISST dataset, applying a 2-year running mean to the data. As in previous figures, the thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a χ^2 distribution for the RMSE.

The extensive skill assessment of the QM decadal predictions has been published in Volpi et al 2021.

List of publications/reports from the project with complete references

Volpi D., Meccia V.L., Guemas V., Ortega P., Bilbao R., Doblas-Reyes F.J., Amaral A., Echevarria P., Mahmood R. and Corti S. (2021a) A Novel Initialization Technique for Decadal Climate Predictions. *Front. Clim.* 3:681127. doi: 10.3389/fclim.2021.681127

Presentations in International conferences and meetings:
- Volpi D.: Updates on the decadal hindcast initialised with quantile matching, EC-Earth Meeting, 10th February 2021, teleconference, (oral presentation).
- Volpi D., Bilbao R., Corti S., Doblas-Reyes F., Guemas V., Meccia V. and Ortega P.: Preliminary results from a quantile matching initialised decadal hindcast, EC-Earth Meeting, 3rd-5th March 2020, teleconference, (oral presentation).
- Volpi D., Meccia V., Ortega P., Guemas V. and Corti S.: An innovative initialisation technique for decadal climate predictions, Fall Meeting 2019, American Geosciences Union, (poster presentation).

Future plans

(Please let us know of any imminent plans regarding a continuation of this research activity, in particular if they are linked to another/new Special Project.)

The findings of this work have opened up the idea of exploiting further the quantile matching technique with the design of a new decadal prediction experiment: the new initialisation would include on the one hand the introduction of the QM for the sea-ice model component. On the other hand, the initial conditions would be taken from a nudging simulation in order to improve the physical consistency between the variables of different components and avoid the potential initial shock. The new experiment would therefore include nudging in the ocean and the sea-ice components and the model would be relaxed towards the quantile matching states.

A new special project is going to be request in the near future in order to implement the ocean and sea-ice quantile matching nudging and the multi-annual predictions associated.

References

Bilbao R, Wild S, Ortega P, Acosta-Navarro J, Arsouze T, Bretonniére P-A, et al. (2021). Assessment of a full-field initialized decadal climate prediction system with the cmip6 version of ec-earth. *Earth Syst. Dyn*. **12**, 173–196. doi: 10.5194/esd-12-173-2021

Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., and Mignot, J. (2021). Improved decadal predictions of north Atlantic subpolar gyre SST in CMIP6. *Geophys. Res. Lett*. 48:e2020GL091307. doi: 10.1029/2020GL091307

Döscher R, Acosta M, Alessandri A, Anthoni P, Arneth A, Arsouze T, et al. (2021): The ec-earth3 earth system model for the climate model intercomparison project 6. *Geoscientific Model Development Discussions* 2021, 1-90. doi:10.5194/gmd-2020-446

Pohlmann H, Müller W A, Kulkarni K, Kameswarrao M, Matei D, Vamborg F S E, et al. (2013): Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophys. Res. Lett*. **40**, 5798–5802. doi: 10.1002/2013GL058051

Smith T, Reynolds R, Peterson T, and Lawrimore J (2008). Improvements to noaa's historical merged land-ocean surface temperature analysis (1880-2006): *J. Clim*. **21**, 2283-2296. doi: 10.1175/2007JCLI2100.1