

Geophysical Research Letters

RESEARCH LETTER

10.1029/2020GL087900

Key Points:

- Development and application of a technique to constrain uninitialized climate projections using initialized decadal predictions
- The constrained ensemble is more skillful than the unconstrained ensemble even after 10 years over the North Atlantic Gyre region
- The potential of the proposed technique crucially depends on the value added by initialization

Supporting Information:

- Supporting Information S1

Correspondence to:

D. J. Befort,
daniel.befort@physics.ox.ac.uk

Citation:

Befort, D. J., O'Reilly, C. H., & Weisheimer, A. (2020). Constraining projections using decadal predictions. *Geophysical Research Letters*, 47, e2020GL087900. <https://doi.org/10.1029/2020GL087900>

Received 25 MAR 2020

Accepted 19 JUL 2020

Accepted article online 21 AUG 2020

Constraining Projections Using Decadal Predictions

Daniel J. Befort¹ , Christopher H. O'Reilly^{1,2} , and Antje Weisheimer^{2,3} 

¹Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK, ²National Centre for Atmospheric Science, University of Oxford, Oxford, UK, ³European Centre for Medium-Range Weather Forecasts, Reading, UK

Abstract There is increasing demand for robust, reliable, and actionable climate information for the next 1 to 50 years. This is challenging for the scientific community as the longest initialized predictions are limited to 10 years (decadal predictions). Thus, to provide seamless information for the upcoming 50 years, information from decadal predictions and uninitialized projections need to be merged. In this study, the ability to obtain valuable climate information beyond decadal time scales by constraining uninitialized projections using decadal predictions is assessed. The application of this framework to surface temperatures over the North Atlantic Subpolar Gyre region, shows that the constrained uninitialized subensemble has higher skill compared to the overall projection ensemble also beyond 10 years when information from decadal predictions is no longer available. Though showing the potential of such a constraining approach to obtain climate information for the near-term future, its utility depends on the added value of initialization.

Plain Language Summary Recently, large efforts have been made toward developing predictions up to 10 years. These predictions are initialized using observations and have shown to be skillful in predicting the observed variability of key variables as, for example, temperature. However, these simulations are not sufficient to meet the increasing demand for seamless climate information up to 50 years. Only climate projections (which are not initialized using observations and are in turn less powerful to predict shorter-term variability) are available to cover these long time scales. To obtain most valuable climate information for the next 1–50 years, this study assesses in how far it is possible to combine information carried by both sources. This is done by choosing for each start date during 1960 and 2005 an ensemble of climate projections, which shows to have the best match with the decadal prediction mean for the next 10 years. Basic skill metrics suggest that it is possible to obtain more skillful predictions using such a framework compared to just using climate projections, even beyond 10 years when no information from decadal predictions are available any longer. However, the value of this methodology heavily depends on the added value of initialization.

1. Introduction

There is an increasing demand for robust, reliable, and actionable climate information for the near-term future (i.e., up to 50 years). With a strong focus on Europe, the EU Horizon 2020 EUCP project aims to deliver such climate information (Hewitt & Lowe, 2018). Climate predictions on these time scales require accurate assimilation of the observed initial state (especially for short lead times up to 10 years) as well as a correct representation of external forcings, for example, greenhouse gas emissions (especially for longer lead times beyond 10 years) (Meehl et al., 2009).

At present information of future climate beyond decadal time scales can only be obtained from uninitialized climate model projections. However, because uninitialized projections represent the climates response to external forcings (e.g., from greenhouse gas concentrations), they are generally not aligned with observed natural variability at a given point in time. Therefore, uninitialized projections are less suitable for near-term future assessments for which large parts of the predictable component stem from an accurate representation of the observed state at initialization. Nonetheless, uninitialized climate projections have been extensively used to derive climate information for the near- and far-term future (Collins et al., 2013; Kirtman et al., 2013). Current state-of-the art climate models differ drastically in their ability to represent observed characteristics of the climate system, which might also limit their capability to simulate changes due to enhanced greenhouse gas forcings. To take the individual model performance into account, several

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

different approaches have been employed to constrain uninitialized projections (see Brunner et al., 2020 and references therein).

In addition to uninitialized climate projections, extensive efforts have been made to develop initialized predictions for decadal time scales (Meehl et al., 2009, 2014). In contrast to uninitialized projections, decadal predictions are initialized using information about the observed climate state and can be superior to projections. Several studies analyzed the added value of initialized predictions over uninitialized climate projections (Doblas-Reyes et al., 2013; Smith et al., 2019). In general, it is found that decadal predictions are mainly superior for the first years after initialization; however, the skill for projections and predictions converges with lead time (Doblas-Reyes et al., 2013). One major exception is over the North Atlantic Subpolar Gyre region for which decadal predictions show added value over projections for lead times up to a decade (Kirtman et al., 2013; Smith et al., 2010).

Based on previous studies, arguably the most valuable climate information for the next 50 years could be obtained by using decadal predictions up to 10 years and climate projections for periods beyond 10 years into the future. For example, by blending both data sets, which has previously been done for short-term forecasts (Kober et al., 2012). However, appending or blending decadal predictions and climate projections is not straightforward, particularly due to the existence of systematic biases in both data sets. For decadal predictions, it is generally necessary to remove a lead-time-dependent bias (e.g., Boer et al., 2016; Goddard et al., 2013). However, biases are not necessarily stationary (Hermanson et al., 2018; Kruschke et al., 2016) and it is not guaranteed that the decadal prediction will seamlessly blend into the climate projection after 10 years.

Here, we propose and test a novel approach which involves constraining uninitialized climate projections using initialized decadal predictions. The framework is motivated by previous studies focusing on shorter time scales, such as that by Ding et al. (2019), who demonstrated the potential to obtain skilful El Niño–Southern Oscillation (ENSO) forecasts on interannual time scales using a model-analog method. In Ding et al. (2019), the subensemble is selected based on the distance of the ensemble members to the observed climate state at a specific point in time. In the constraining method presented in this study, the uninitialized ensemble is constrained using decadal predictions instead of observations and the selection of the individual model integrations is performed by taking into account several years rather than just one certain point in time. The constrained ensemble solely consists of uninitialized projections, which means that this approach bypasses the potential problems arising when blending or appending both data sets directly.

In this study, the proposed constraining method is applied to regional surface air temperatures using CMIP5 simulations. The data used are outlined in section 2; section 3 explains the methods used to constrain projections using predictions. Results applying this method to CMIP5 surface air temperature data are presented in section 4, followed by a discussion and conclusions in section 5.

2. Data and Methods

The backbone of this study is annually averaged surface air temperature from uninitialized and initialized CMIP5 models and observations for the period from 1961 until 2015. The HadCRUT4.6.0.0 ensemble median is used for observations (Morice et al., 2012). Nine decadal prediction models with a total of 72 members and 44 uninitialized projection models with a total of 154 members are used (Table S1; supporting information). The baseline period used to bias correct the simulation data spans from 1970 to 2006, inclusive. This period is chosen as it allows for the removal of the same climatology for each lead time in the decadal predictions. For uninitialized projections and observations, the mean over the baseline period is removed. For each decadal prediction model, the lead time dependent bias is calculated using only those forecast years within the baseline period. All decadal predictions used in this study are initialized for each year from 1960 until 2005. Three specific regions are analyzed in this study: North Atlantic Subpolar Gyre region (GYRE), [45–60°N; 50–20°W], Europe [35–71°N; 10°W to 40°E], and NINO3.4 [5°S to 5°N; 170–120°W]. The skill of the different ensembles is assessed using the anomaly correlation coefficient (ACC) and root-mean-square error (RMSE) metrics. To allow the assessment of lead-time-dependent skill up to 15 forecast years using observational data until 2015, only start dates between 1960 and 2000 are used from the decadal predictions. It should be noted that HadCRUT data are blended products of CRUTEM4 land surface air temperatures

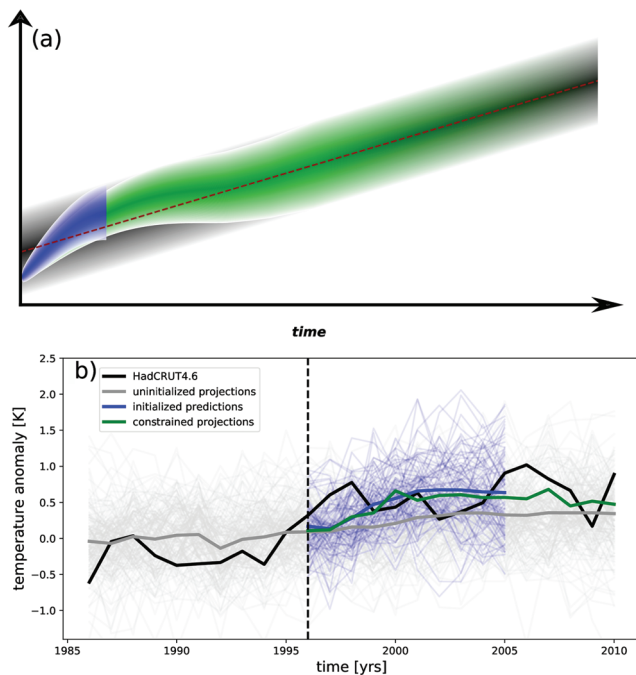


Figure 1. (a) Schematic illustration of framework used to constrain uninitialized projections (gray) using decadal predictions (blue). The selected ensemble based on proximity to the decadal prediction ensemble mean are indicated in green. Shaded areas indicate range of the respective ensemble. (b) Example for constrained uninitialized subensemble (green) using decadal predictions ensemble mean (blue) for surface temperatures over the North Atlantic Gyre region. Unconstrained projection ensemble mean is shown in gray and observations (HadCRUT4.6) in black. Ensemble mean values are shown as thick lines, whereas thinner lines show individual ensemble members (unconstrained projections and decadal predictions only).

and HadSST3 sea surface temperatures, whereas we use surface air temperature from the model simulations for comparison.

3. Framework for Constraining Uninitialized Projections

The main difference between predictions and projections is that predictions are assumed to be more closely aligned to the observed phases of natural variability. This is due to the assimilation of the observed climate state into the initialization, which is not the case for uninitialized projections. The aim of this study is to examine to what extent the additional information carried by decadal predictions can be utilized to usefully constrain an ensemble of uninitialized projections. A schematic demonstrating how initialized predictions could be used to constrain an ensemble of uninitialized projections for a specific point in time is shown in Figure 1a. In this schematic example, the uninitialized projections (gray) are well able to simulate the signal linked to external forcings; however, its spread is quite large due to the fact that the underlying ensemble is not aligned with the observed mode of natural variability. In contrast, the spread of the initialized prediction (blue) is much smaller, especially for short lead times. Furthermore, in this schematic example, the ensemble mean of the predictions is different to the ensemble mean of the projections, indicating a stronger positive signal over the upcoming decade. If we assume that the decadal prediction provides a more skillful prediction of the upcoming decade than the uninitialized projections, it is sensible to constrain the projections based on the prediction ensemble mean. The constraining is performed by subselecting, from all uninitialized projections (gray), those ensemble members which are closest to the decadal prediction ensemble mean over the following decade. This constrained uninitialized ensemble (green) can be used not only for the upcoming decade but also beyond 10 years, when decadal predictions are unavailable.

4. Results

4.1. Decadal Predictions Versus Projections—Added Value

One of the requirements for the proposed constraining framework to be effective is the existence of higher skill levels in the decadal predictions compared to uninitialized projections. In the constraining framework, those model integrations that best match the mean of the decadal prediction are selected. Therefore, if the decadal ensemble mean is not skillful, the constrained projection ensemble will likely also not be skillful. Several studies have assessed the added value of CMIP5 predictions over projections for different variables, regions, and lead times. One of the key regions for which surface temperatures are improved, even beyond several years, is the North Atlantic Subpolar Gyre (GYRE) region. Figure 2 shows ACC and RMSE for uninitialized and initialized simulations, using HadCRUT as reference for surface air temperatures over the GYRE. ACC is between 0.6 and 0.8 for decadal predictions over the first 10 forecast years. In contrast, ACC for uninitialized projections increases with forecast time from about 0.2 to about 0.8. The variable skill in time for uninitialized projections seems counterintuitive but is caused by the different time periods used for the verification of the different forecast lead times. For example, the years 1961 until 2001 are used for forecast year 1, whereas the years 1970 to 2010 are used for forecast Year 10. If the analysis is restricted to the years 1970–2006 (which are available at all lead times for the decadal forecasts), the ACC/RMSE values for projections are constant over the first 10 forecast years (as shown in Figures S1a and S1b). In contrast, for decadal predictions the ACC values decrease with lead time and RMSE values increase with lead time (Figures S1a and S1b). Furthermore, high correlations are partly related to the observed warming trend over the GYRE region during the twentieth century, which is, to a large extent, captured by predictions and projections. Removing this trend (e.g., by linearly detrending) leads to smaller correlations for predictions and projections (Figures S1c and S1d). However, using only the years from 1970 to 2006 reduces the sample size,

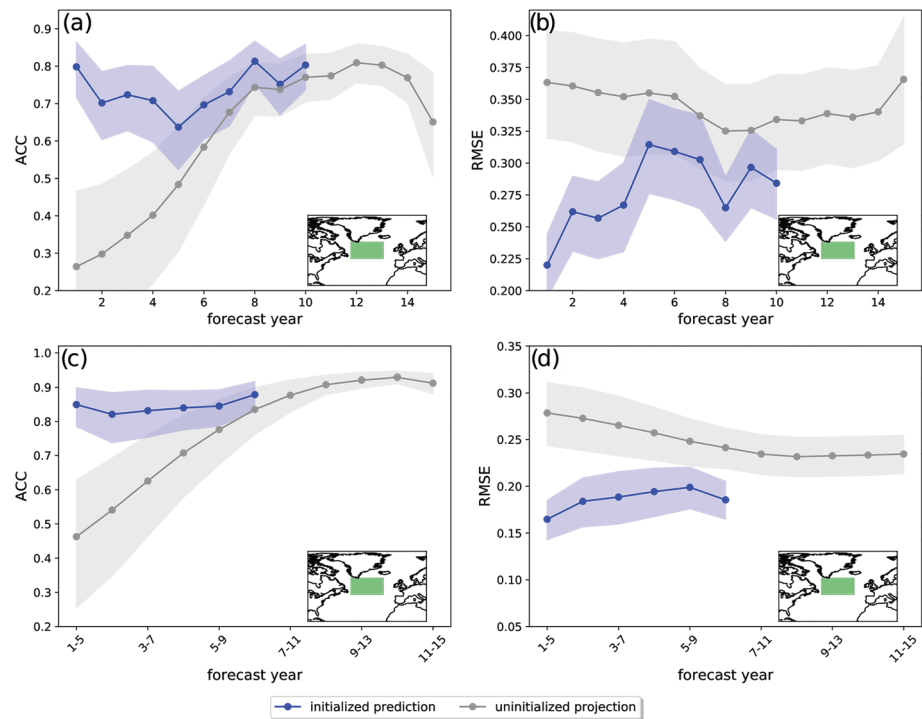


Figure 2. (a) Anomaly correlation coefficient (ACC) between observations and uninitialized projection (gray)/decadal predictions (blue) for surface temperatures over the North Atlantic Gyre region for a given forecast year. Shaded areas show 10–90% confidence intervals (based on a 10,000 sample bootstrap). (b) Same for root-mean-square error (RMSE). (c) and (d) same as (a) and (b) but for 5-year averages.

which is why we use all available years at specific lead times hereafter, even though this results in an increased skill with lead time. RMSE are lowest for early forecast years for the decadal predictions, whereas relatively constant values are found for uninitialized projections. Again, decadal predictions are more skillful than projections in terms of RMSE, which is consistent with the higher ACC skill but even more pronounced. The higher skill in the decadal predictions is better visible for 5-year averages compared to the annual values, likely because the contribution of unpredictable interannual variability is reduced in the decadal predictions (Figures 2c and 2d). It should also be noted that skill over the GYRE region in uninitialized projections might arise from the incorrect representation of physical mechanisms in these simulations as discussed by Yeager et al. (2012).

The same analysis is carried out for the regions covering Europe and the central tropical Pacific NINO3.4 region (Figures S2 and S3). For both regions, the added value of initialization is smaller compared to the GYRE region. For NINO3.4, decadal predictions are more skillful for lead times of up to 2 years, which is in line with current studies on the predictability of ENSO on multiyear time scales (Dunstone et al., 2016). For Europe, the added value of decadal predictions over projections is restricted to the first forecast year, which seems surprising given that there is added value over parts of the North Atlantic in decadal predictions for longer lead times. However, it has been shown that current models do have limitations in representing some of the observed teleconnections from the North Atlantic (e.g., Simpson et al., 2018, 2019). Furthermore, in the presence of strong externally forced trends the added value of initialized forecasts can appear small if measured by differences in ACC and it has been proposed to measure the added value based on the residual correlation after regressing out the uninitialized ensemble mean signal from the initialized predictions and observations instead (Smith et al., 2019).

4.2. Constraining Temperatures Over the North Atlantic Subpolar Gyre Using Decadal Predictions

As shown in the previous section, decadal predictions demonstrate higher levels of skill for GYRE surface temperatures compared with uninitialized projections (in terms of higher correlations and lower RMSEs)

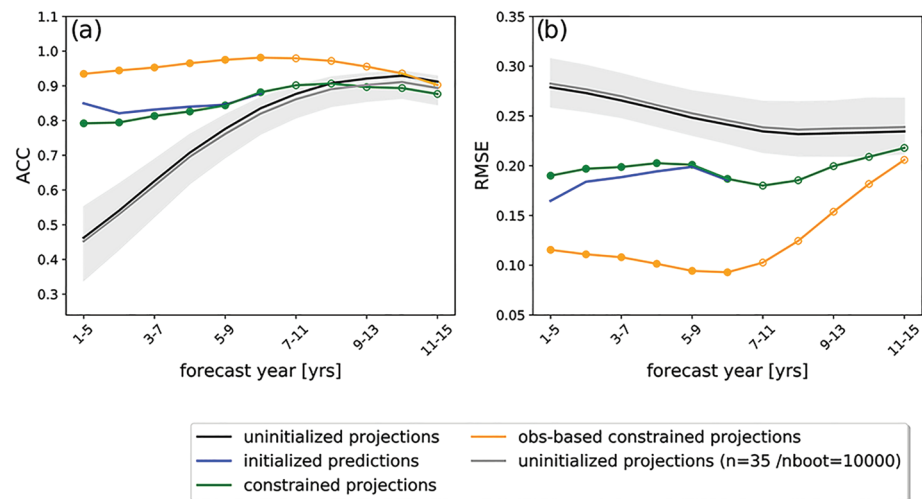


Figure 3. (a) ACC and (b) RMSE for unconstrained projections (black), constrained projections using decadal ensemble mean (green), constrained projections using observations (orange), unconstrained projections with fixed ensemble size $n = 35$ (gray) and initialized predictions (blue) for annual mean surface air temperatures over the North Atlantic Gyre. Statistics are given for 5-year averages. The gray area indicates the 10–90% confidence intervals (based on a 10,000 sample bootstrap). Closed circles indicate those periods for which all forecast years have been used to constrain the respective ensemble, whereas open circles indicate those periods for which at least 1 year has not been used for constraining.

even at the end of the prediction. For this reason, the GYRE region is ideal for testing the constraining framework proposed in section 3.

Here, for each individual start year from 1960 to 2000, we chose the 35 uninitialized projection ensemble members with the highest similarity to the respective decadal prediction ensemble mean over the following 10 years. We assessed the sensitivity to the exact ensemble size of the constrained ensemble and the main conclusions are similar for values of 25 and 45 (Figure S4). We also analyzed in how far results are affected when using fewer years to constrain the projections, for example, only forecast Years 1 to 5 from the prediction, which in general leads to smaller skill of the constrained ensemble (not shown). The similarity of an individual uninitialized projection to the decadal ensemble mean is measured by the mean absolute error (MAE) of all 10 annual mean temperature values for the following decade. To select those uninitialized projections, which simulate long-term variability changes similar to the decadal prediction ensemble mean, all uninitialized projection members are smoothed using a Hanning window with a window length of 11 years. This smoothing intends to remove the interannual variability of each member, which is done implicitly for the decadal predictions ensemble mean by averaging over all 72 members.

The resulting constrained ensemble using the decadal forecast initialized in 1995 is shown in Figure 1b. The ensemble mean from the decadal prediction follows the observations more closely than the uninitialized ensemble mean, which is reflected in a 15% smaller MAE for decadal predictions compared to uninitialized projections for the decade 1996–2005. The constrained 35-member ensemble follows the ensemble mean of the predictions very closely, since they are those with lowest average absolute error with respect to the decadal ensemble mean. This results in a nearly 15% smaller MAE for the constrained compared to the unconstrained projections over the following 10 years (1996–2005). In addition to the closer agreement with observations during the first decade, the constrained ensemble mean also shows a higher agreement at forecast times of 11–15 years (2006–2010), which is reflected in a reduction of the MAE of 28% compared to the whole uninitialized ensemble. For this example, the MAE of the constrained ensemble over the whole 15-year period is reduced by 20% compared to the unconstrained projections.

To fully examine the efficacy of the constraining method, the subselection of the uninitialized projections is conducted for all start dates of decadal predictions within the period 1960 to 2000, inclusive. Using this period, together with observational data, the added value of the selected subensemble over the overall uninitialized ensemble can be assessed. ACC and RMSE against forecast time is shown in Figure 3. Predictions on interannual time scales cannot be expected to capture the observed year to year variability, so statistics

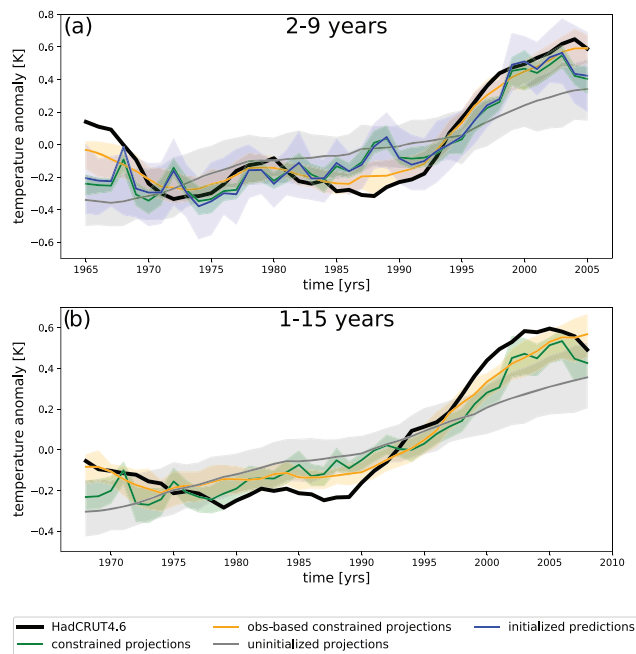


Figure 4. Surface air temperature anomalies over North Atlantic Gyre region (a) averaged for forecast Years 2 to 9, and (b) averaged over forecast Years 1 to 15. Uninitialized projections are shown in gray, initialized predictions in blue (only for a), the constrained projections using decadal ensemble mean in green, the constrained projections using smoothed observations in orange, and observations in black. The shaded areas indicate the interquartile range of the respective ensemble.

(Figure 4a). The benefits of constraining become even more remarkable during the last decade of the twentieth century, when the strong warming trend is not captured by the uninitialized (unconstrained) projection but is well represented by the constrained ensemble and decadal prediction system. As the decadal prediction outperforms the constrained ensemble over the first 10 forecast years, it is of greater interest to analyze lead times beyond 10 years. The evolution of the 1 to 15 year averaged ensembles confirm that the constrained subensemble shows a better agreement with observations than the overall unconstrained projections, particularly during the period at the end of the twentieth century (Figure 4b). It should be noted that even though showing a similar ensemble mean, the constrained ensemble has a smaller ensemble spread compared to the decadal prediction ensemble (Figure 4a). We find that both ensembles are not reliable, with the decadal prediction being underconfident whereas the constrained ensemble is overconfident for the first 10 forecast years (see supporting information Figure S6 for further information).

Using the same technique projections of surface temperature are constrained for Europe and the NINO3.4 region (Figure S5). The skill of the constrained ensemble is smaller over both regions compared to the GYRE region, which is most probably related to the smaller added value of the decadal predictions over projections over Europe/NINO3.4 (Figures S2 and S3) compared with the GYRE region (Figure 2).

4.3. Potential Upper Skill Limits of Constrained Projections

The potential level of skill increase of the constrained ensemble is limited by the added value of the decadal prediction over the ensemble of uninitialized projections. The implications of this are clearly visible in the time series averaged for lead times of 2–9 years (Figure 4a). While the constrained projections using decadal predictions (green) clearly show an improved ability to capture the observed variability compared to unconstrained projections, they are still not able to simulate observed variability during some periods. One such example is the period during the late 1980s, with cooler conditions over the GYRE region, which is not captured by the constrained ensemble. This can be explained by the fact that the decadal prediction is also not able to capture these lower temperatures during that decade. This raises the question of how much better a

are calculated over consecutive 5-year intervals. ACC values for the unconstrained projections increase with lead time, as already shown in Figure 2, whereas ACC values for the decadal prediction ensemble are relatively stable over lead time. ACC values for the constrained subensemble are similar to those found for the decadal forecast ensemble over the first 10 years. From year 11 to 15 ACC values for the constrained ensemble are about the same as found for the unconstrained projections. ACC is known to be sensitive to the ensemble size, thus, the constrained (35 members) ensemble should rather be compared to an unconstrained ensemble with the same size. It is found that ACC skill is not significantly different between a randomly drawn 35-member ensemble and the constrained subensemble. Improvements are more noticeable for RMSE. The constrained ensemble has a RMSE which is slightly higher than the decadal predictions for the first 10 years but smaller than the unconstrained full ensemble. As well as exhibiting smaller RMSE than the unconstrained full ensemble during the first 10 years, the RMSE is also significantly smaller in Years 11–15, a period when the decadal predictions are no longer available. Averaged over the 15 forecast years, RMSE is reduced by as much as 39% compared to the uninitialized ensemble. For the last 5 years beyond the decadal prediction horizon (Years 11–15), error reduction is still about 7%.

The temporal evolution of the constrained ensembles averaged for forecast years 2 to 9 and 1 to 15 are shown in Figure 4. For forecast Years 2 to 9, the constrained subensemble closely corresponds to the decadal forecast ensemble mean, which is expected as the latter was used during the subselection process. To some extent both the decadal prediction and constrained projections capture the warmer conditions around the mid 1960s, which is not captured by the uninitialized unconstrained projections

constrained subensemble can be if using a perfectly skilful forecast for selecting uninitialized projections. The upper limit can be tested by using observations, rather than the decadal forecast ensemble average, to constrain the uninitialized projections. Note that as this method selects those uninitialized ensemble members closest to observations for the following decade, it is clearly not applicable for real-time predictions. To focus on decadal-scale variability, interannual variability is removed from the observations in the same way as it was done for the uninitialized projections. The resulting time series averaged over lead years 2 to 9 and 1 to 15 are shown in Figures 4a and 4b. The observationally constrained projections capture the observed variability much better than decadal predictions or the prediction-based constrained ensemble (Figure 4a). During the 1960s and late 1980s in particular, the observation-based constrained ensemble is able to reproduce the observed temperature anomalies much better than the prediction-based constrained ensemble. The observation-based constrained ensemble has a higher ACC over the entire 15 forecast years compared to the prediction-based constrained ensemble (Figure 3a), even though the difference is small for long lead times. Differences are more notable for RMSE, which is reduced significantly (more than 60%) compared to the overall uninitialized projections for the first 10 years (Figure 4b). This improvement is about double the increase in skill seen in the prediction-based constrained ensemble (Figure 3b). For longer lead times, beyond 12 years, differences between both of the constrained ensembles decrease, suggesting that even a perfect decadal forecast would be unable to provide further added skill in this framework.

5. Summary and Discussion

To obtain seamless information for the upcoming 50 years, information from initialized predictions (which are limited to 10 years) and uninitialized projections (available until the end of the 21st century) need to be combined. Here, we have presented a novel technique to fill this gap by constraining uninitialized climate projections using decadal predictions. The proposed framework has been applied to data from the CMIP5 archive for surface air temperatures over the North Atlantic Subpolar Gyre region (GYRE). The GYRE region has been chosen because decadal predictions are found to have higher skill compared to uninitialized projections even for long lead times. Constraining the projections using predictions can only be valuable if the decadal predictions carry additional information over the uninitialized ensemble.

Results suggest that uninitialized ensembles constrained in such a way have similar performance compared to unconstrained uninitialized ensembles if measured by correlation score. However, the constrained subensemble has smaller forecast errors (RMSE), not only over the first 10 forecast years for which decadal predictions have been used as a constraint but also for lead times beyond 10 years. These results show that it is possible to obtain a subselected ensemble, from the uninitialized projections, which is more skillful than using the whole ensemble of projections. It is further shown that the entire observed variability is not captured by the constrained ensemble, which is partly because the decadal predictions themselves do not capture this observed variability. To test the potential upper limit of using the proposed framework for a perfect decadal prediction system, observations are used to constrain the uninitialized climate projections. As expected, this observation-based constrained ensemble performs better than the subensemble constrained using decadal predictions. However, after about 15 years it seems that the level of skill is not distinguishable between the observation-based and prediction-based constrained subensembles, which might suggest that the external forcing begins to dominate over the internal variability. However, the observationally constrained ensemble does not reflect a definite upper limit as constraining might be potentially superior if using other remote regions and/or variables to constrain GYRE sea surface temperatures.

There are several factors limiting the potential added value using this framework. First, the decadal prediction ensemble must be at least as skillful as the ensemble of projections. This is obvious as the selection of the uninitialized projections is based on their proximity to the decadal prediction mean. Therefore, for the subsampled ensemble to have higher skill than the unconstrained projections, the decadal prediction ensemble mean needs to provide additional information. Second, in order for the constrained ensemble to have added value compared with the full projections beyond the following decade (when decadal predictions are no longer available), the climate system needs to exhibit variability on decadal or longer time scales as the subselecting is able to primarily select only those projections which are aligned with the observed decadal phase. This need for decadal variability motivates the application of this framework to oceanic variables related to the large-scale ocean circulation. In this study the method has been applied to surface temperatures because

these are available for all CMIP5 simulations. However, in future studies it is important to assess the potential additional value of constraining uninitialized projections by using oceanic variables (e.g., ocean heat content), especially as recent studies indicate that North Atlantic temperature predictability is related to anomalous ocean heat transport (e.g., Borchert et al., 2018, 2019).

A further limitation of this study is the small sample size of 41 initial dates (1960 until 2000), which are available to evaluate the method. The robustness of the proposed methodology, as well as investigation of further improvements, could be tested thoroughly using a perfect model framework. One advantage of such perfect model approaches is that they do not suffer from lead-time-dependent biases. While these lead time dependent biases are often assumed to be stationary, this is not necessarily the case (e.g., Hermanson et al., 2018; Kruschke et al., 2016). Besides the methods dependency on the ability of the initialized predictions to capture observed climate variability, it is also crucial that the observations lie within the uninitialized ensemble.

Overall, this study demonstrates an easy-to-apply method to obtain skillful climate information beyond decadal time scales by constraining uninitialized projections using decadal predictions. However, given the strong dependence of the proposed framework on the added value of initialized decadal predictions over noninitialized projections, further improvements of the current decadal prediction systems are necessary so that the proposed method can be successfully applied to more regions and variables.

Data Availability Statement

In this study uninitialized historical, RCP4.5 and RCP8.5 simulation data as well as the CMIP5 decadal hindcasts listed in Table S1 (supporting information) can be retrieved via the ESGF (<https://esgf-node.llnl.gov/search/cmip5/>). The MPI-ESM-LR and MPI-ESM-HR decadal hindcasts from the SPECS project are also available on the ESGF (<https://esgf-node.llnl.gov/search/esgf-llnl/>).

Acknowledgments

Daniel J. Befort, Christopher H. O'Reilly, and Antje Weisheimer received funding from the European Union under Horizon 2020 (Grant Agreement 776613). The authors like to thank Francisco Doblas-Reyes and an anonymous reviewer for their helpful comments.

References

- Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9(10), 3751–3777. <https://doi.org/10.5194/gmd-9-3751-2016>
- Borchert, L. F., Düsterhus, A., Brune, S., Müller, W. A., & Baehr, J. (2019). Forecast-oriented assessment of decadal hindcast skill for North Atlantic SST. *Geophysical Research Letters*, 46, 11,444–11,454. <https://doi.org/10.1029/2019GL084758>
- Borchert, L. F., Müller, W. A., & Baehr, J. (2018). Atlantic Ocean heat transport influences interannual-to-decadal surface temperature predictability in the North Atlantic region. *Journal of Climate*, 31(17), 6763–6782. <https://doi.org/10.1175/JCLI-D-17-0734.1>
- Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., et al. (2020). Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework. *Journal of Climate*, 33(20), 8671–8692. <https://doi.org/10.1175/jcli-d-19-0953.1>
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In *Climate change 2013—The physical science basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). United Kingdom and New York, NY, USA: Cambridge University Press.
- Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2019). Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophysical Research Letters*, 46, 1721–1730. <https://doi.org/10.1029/2018GL080598>
- Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., Garca-Serrano, J., Guemas, V., Kimoto, M., et al. (2013). Initialized near-term regional climate change prediction. *Nature Communications*, 4, 1715. <https://doi.org/10.1038/ncomms2704>
- Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., et al. (2016). Skillful predictions of the winter North Atlantic Oscillation one year ahead. *Nature Geoscience*, 9, 809–814. <https://doi.org/10.1038/ngeo2824>
- Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., et al. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40, 245–272. <https://doi.org/10.1007/s00382-012-1481-2>
- Hermanson, L., Ren, H.-L., Vellinga, M., Dunstone, N. D., Hyder, P., Ineson, S., et al. (2018). Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Climate Dynamics*, 51, 1411–1426. <https://doi.org/10.1007/s00382-017-3962-9>
- Hewitt, C. D., & Lowe, J. A. (2018). Toward a European climate prediction system. *Bulletin of the American Meteorological Society*, 99(10), 1997–2001. <https://doi.org/10.1175/BAMS-D-18-0022.1>
- Kirtman, B., Power, S. B., Adedoyin, A. J., Boer, G. J., Bojariu, R., Camilloni, I., et al. (2013). Near-term climate change: Projections and predictability. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Kober, K., Craig, G. C., Keil, C., & Dörnbrack, A. (2012). Blending a probabilistic nowcasting method with a high-resolution numerical weather prediction ensemble for convective precipitation forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138, 755–768. <https://doi.org/10.1002/qj.939>
- Kruschke, T., Rust, H. W., Kadow, C., Müller, W. A., Pohlmann, H., Leckebusch, G. C., & Ulbrich, U. (2016). Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms. *Meteorologische Zeitschrift*, 25(6), 721–738. <https://doi.org/10.1127/metz/2015/0641>
- Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., et al. (2014). Decadal climate prediction: An update from the trenches. *Bulletin of the American Meteorological Society*, 95(2), 243–267. <https://doi.org/10.1175/BAMS-D-12-00241.1>

- Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction: Can It Be Skillful? *Bulletin of the American Meteorological Society*, 90(10), 1467–1486. <https://doi.org/10.1175/2009BAMS2778.1>
- Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, 117, D08101. <https://doi.org/10.1029/2011JD017187>
- Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018). Modeled and Observed Multidecadal Variability in the North Atlantic jet stream and its connection to sea surface temperatures. *Journal of Climate*, 31(20), 8313–8338. <https://doi.org/10.1175/JCLI-D-18-0168.1>
- Simpson, I. R., Yeager, S. G., McKinnon, K. A., & Deser, C. (2019). Decadal predictability of late winter precipitation in western Europe through an ocean-jet stream connection. *Nature Geoscience*, 12, 613–619. <https://doi.org/10.1038/s41561-019-0391-x>
- Smith, D. M., Eade, R., Dunstone, N. J., Fereday, D., Murphy, J. M., Pohlmann, H., & Scaife, A. A. (2010). Skilful multi-year predictions of Atlantic hurricane frequency. *Nature Geoscience*, 3, 846–849. <https://doi.org/10.1038/ngeo1004>
- Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., et al. (2019). Robust skill of decadal climate predictions. *npj Climate and Atmospheric Science*, 2, 13. <https://doi.org/10.1038/s41612-019-0071-y>
- Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal prediction case study: Late Twentieth-century North Atlantic Ocean heat content. *Journal of Climate*, 25(15), 5173–5189. <https://doi.org/10.1175/JCLI-D-11-00595.1>