

Geophysical Research Letters



RESEARCH LETTER

10.1029/2019GL084223

Key Points:

- A model for the behavior of initialized decadal predictions is presented and then used to develop a technique for postprocessing predictions
- The technique is shown to improve hindcast skill in scenarios with widely different bias and interannual variability characteristics
- The technique throws light on issues such as expected versus realized time scale for loss of memory of IC and regional quirks in initialization

Supporting Information:

- Supporting Information S1

Correspondence to:

B. T. Nadiga,
 balu@lanl.gov

Citation:

Nadiga, B. T., Verma, T., Weijer, W., & Urban, N. M. (2019). Enhancing skill of initialized decadal predictions using a dynamic model of drift. *Geophysical Research Letters*, 46, 9991–9999. <https://doi.org/10.1029/2019GL084223>

Received 26 JUN 2019

Accepted 12 AUG 2019

Accepted article online 16 AUG 2019

Published online 31 AUG 2019

Corrected 23 OCT 2019

This article was corrected on 23 OCT 2019. See the end of the full text for details.

Published 2019. This article is a U.S. Government work and is in the public domain in the USA.

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.

Enhancing Skill of Initialized Decadal Predictions Using a Dynamic Model of Drift

Balasubramanya T. Nadiga¹ , Tarun Verma¹ , Wilbert Weijer¹ , and Nathan M. Urban¹ 

¹Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract Since near-term predictions of present-day climate are controlled by both initial condition predictability and boundary condition predictability, initialized prediction experiments aim to augment the external-forcing-related predictability realized in uninitialized projections with initial-condition-related predictability by appropriate observation-based initialization. However, and notwithstanding the considerable effort expended in finding such “good” initial states, a striking feature of current, state-of-the-art, initialized decadal hindcasts is their tendency to quickly drift away from the initialized state, with attendant loss of skill. We derive a dynamical model for such drift, and after validating it we show that including a recalibrated version of the model in a postprocessing framework is able to significantly augment the skill of initialized predictions beyond that achieved by a use of current techniques of postprocessing alone. We also show that the new methodology provides further crucial insights into issues related to initialized predictions.

1. Introduction

The ability to predict near-term changes in climate, here understood as up to a few decades, is as scientifically challenging (e.g., see Meehl et al., 2009, 2014, and references therein) as it is important. The importance of such an ability stems, for example, from the fact that our attempts to assess societal and ecological impacts of climate change and propose, study, and plan adaptation and mitigation strategies are only as accurate and useful as the climate change predictions are correct (e.g., see MacLeod et al., 2012). As such, a series of coordinated decadal prediction experiments have been put in place to help improve understanding of the various issues related to decadal predictions (e.g., see Taylor et al., 2012 and Eyring et al., 2016).

Since near-term variability of present-day climate is influenced by both natural variability and external forcing (Lorenz, 1975), two types of such decadal prediction experiments are typically considered: uninitialized projections (UP) that account for changes in climate due to changes in external forcing and initialized decadal prediction experiments which aim to augment the external-forcing-related predictability realized in UP with natural-variability-related predictability by appropriate observation-based initialization. While uninitialized climate change simulations date back to the earliest attempts at climate modeling, interest in initialized (decadal) predictions (IP) is more recent. Indeed, interest in the latter became mainstream after initial studies such as Smith et al. (2007), Keenlyside et al. (2008), and Pohlmann et al. (2009), about a decade ago, suggested that the predictive aspects of simulations of climate over the past half a century could benefit from observation-based initialization. In the context of such IPs, a primary objective is the improvement in skill of such predictions, beyond that achieved in UPs.

A comparison of predictions of past climate (hindcast) to observations (verification data) typically reveals biases in the former (bottom panel of Figure 1). Such “model bias” can be attributed, on the one hand, to errors in initial conditions (ICs) and, on the other, to “model error.” Here, by model error we refer to deficiencies in the model such as insufficient resolution and inaccurate parameterizations of unresolved processes which lead to the model’s inability to accurately simulate the delicate dynamical balance of processes that underlies both the mean state and the modes of variability of the climate system. It is easy to see how IP efforts can be undermined by model bias: If model bias were to be large in comparison to interannual climate variability (size of the signal to be predicted), the skill in such predictions would necessarily be small (bottom panel of Figure 1). As such, reducing model bias is a top priority of modeling centers (e.g., see Hawkins et al., 2014; Hurrell et al., 2009). But it is also a very challenging task. Consequently, while it is an area of active, ongoing research, progress is typically incremental and slow but steady (e.g., see Drews et al., 2015).

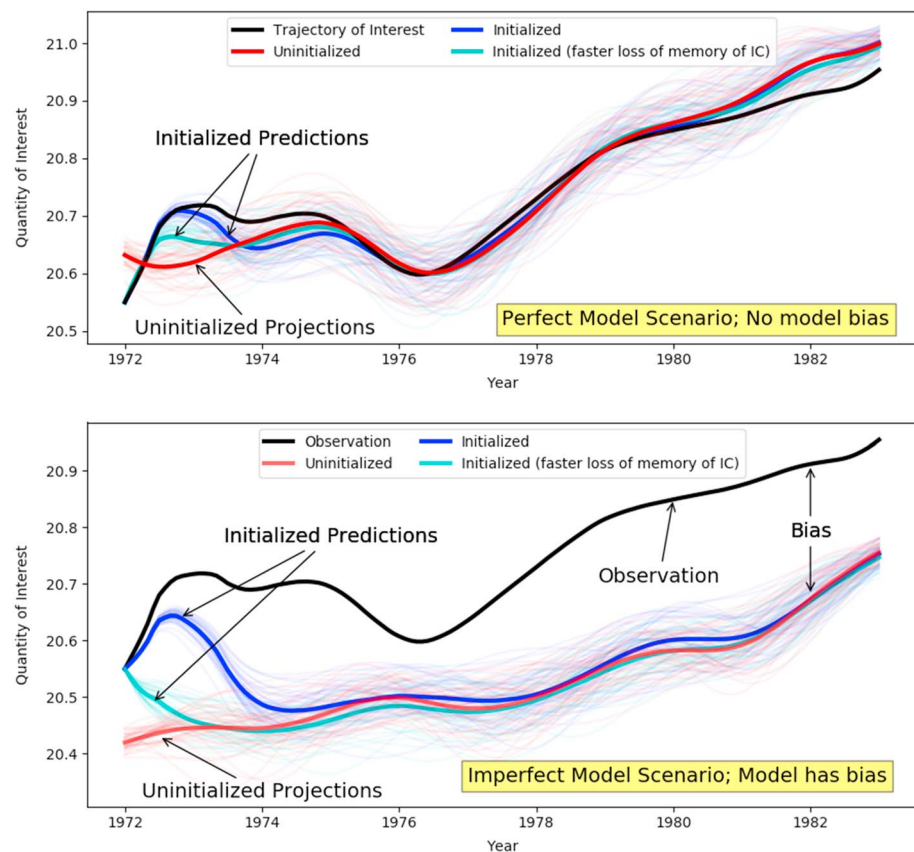


Figure 1. Schematics of initialized (blue and cyan) and uninitialized projections (red) in perfect model (top) and imperfect model (bottom) scenarios. The observation is illustrated in black. The main question we address in this article is the following: Given the IPs (initialized predictions) in cyan, how can they be postprocessed to be more predictive of the observations in black? Answering this question leads to being able to predict near-term changes in climate better. The development of a better postprocessing method has a greater impact in the imperfect model scenario (bottom panel); Coupled Model Intercomparison Project models fall in this category. Our answer is based on first developing a dynamic model that describes the dominant behavior of the IPs (cyan and blue). After validating the model, we show that the new postprocessing method better predicts observations over the hindcast period, when compared to state-of-the-art methods for postprocessing alone. The drift model developed suggests that a faster loss of memory of ICs (initial conditions), as in the cyan IPs, is more typical (cf. IPs in blue) and provides other insights (discussed in the final section).

Given this state of affairs, for current, state-of-the-art IPs to be useful, bias correction in a postprocessing fashion is essential (e.g., see Choudhury et al., 2017; Doblas-Reyes et al., 2013; Kharin et al., 2012).

In IPs, a wide range of techniques are used for either full-field or anomaly initialization and to generate the associated ensemble of initial perturbations (e.g., see Hazeleger et al., 2013; Magnusson et al., 2013; Polkova et al., 2019; Smith et al., 2013). Such techniques range from data assimilation in the ocean component alone (e.g., Pohlmann et al., 2009; Nadiga et al., 2013) to coupled data assimilation (e.g., Yang et al., 2013) to partially coupled spin-up (Thoma et al., 2015), to spinning up ocean and sea ice components with reanalysis products (e.g., Yeager et al., 2018) and others (e.g., see Matei et al., 2012). Even though this is an area of active research, a salient aspect of IPs is that the initial state in these simulations, at a particular start date, is closer to observations than the state of the (possibly an ensemble of) UPs at that same date.

Notwithstanding the considerable effort expended in finding such “good” initial states, a striking feature of current, state-of-the-art, initialized decadal hindcasts is their tendency to drift away from the initialized state (e.g., Hazeleger et al., 2013; Magnusson et al., 2013; Meehl et al., 2014; Smith et al., 2013; bottom panel of Figure 1; Figures 2 and 3; Figure 1 of Kim et al., 2012, and others). Since model drift and model bias are sometimes used interchangeably, for definiteness, we will refer to the early behavior of IPs that leads to a (sometimes quick) departure from its initialized state as “initialization-induced drift” or “initialization drift”

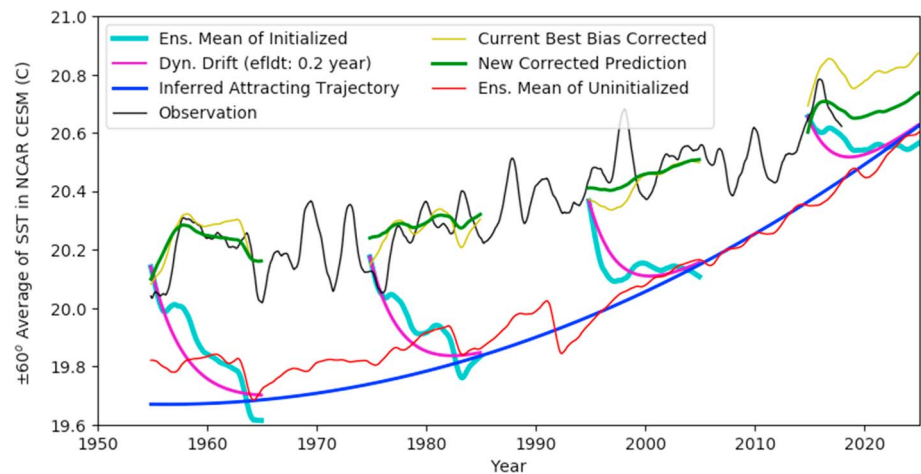


Figure 2. Evolution of the ensemble mean of IPs started at four different times is shown (cyan downward trending lines). When bias corrected using TBC, the resulting predictions $I_{jr}^{(C)}$ are shown in yellow. Drift computed using the dynamical drift model described in section 2 and supporting information is shown in magenta and captures the dominant behavior of the IPs. Also resulting from the model is the attracting trajectory shown in blue. This is seen to follow the uninitialized projection trajectory shown in red; the correspondence is even greater with RDM2 (see Figure S3). These elements are recombined statistically and bias corrected using TBC to obtain the corrected predictions $I_{jr}^{(RC)}$ which are shown in green. The new procedure improves hindcast skill of the IPs: averaged over all lead times and start years, not only is root-mean-square error reduced by 10–15%, but anomaly correlation coefficient increases by about 4–6% (see Tables 1 and S1). IP = initialized predictions; TBC = trend-based bias correction; SST = surface temperature; NCAR = National Center for Atmospheric Research; CESM = Community Earth System Model.

or simply drift. A primary reason for such drift is related to the fact that the climate system loses memory of its IC and transitions from a regime of IC-related (equivalently natural-variability-related) predictability to external forcing-related predictability in near-term predictions. When such a transition occurs in a model that is biased, initialization drift is exacerbated and the IP trajectories appear to almost jump away from the initialized state (bottom panel of Figure 1). Other reasons may include different bias characteristics in the equilibrated states of the system used to obtain the initial state and those used for prediction.

Given that fixing model bias and developing initialization methods are already tasks that are actively worked on and are of high priority at modeling centers, currently, we concern ourselves with the more pragmatic question of whether one can profitably take into account initialization drift in a postprocessing framework to enhance the skill of IPs. Toward this end, we develop a dynamical model for initialization drift and show that a postprocessing methodology based on this model of drift is able to significantly enhance the skill of IPs when compared to using existing state-of-the-art techniques of bias correction alone. We also show that the new methodology provides further crucial insights into issues related to IPs of climate.

In order to illustrate each of the points discussed above, motivate the new model of drift and the postprocessing method presented, and facilitate further related discussions, we start with two schematics in Figure 1. In both of these panels, one is interested in predicting the black trajectory. In each of these two panels, an uninitialized-projection ensemble is shown in red, and two IP ensembles are shown in blue and cyan. The individual ensemble members are shown in thin lines, and the ensemble mean is shown in thick lines. The top panel corresponds to a perfect model scenario, whereas the bottom panel corresponds to an imperfect model scenario and is more realistic.

In each of the panels, the uninitialized projections are seen to capture the upward trend of the black trajectory. The upward trend of the black trajectory itself should be thought of as caused by external forcing. As such, the UPs are seen to capture predictability that arises from external forcing. Next, the schematics illustrate the point that sophisticated state estimation procedures have permitted a proper initialization of the IPs whereby the IPs are coincident with observations at the start of 1972. In the initial period up to about 2 years later, the IP ensembles are seen to be closer to the observations than the UPs are to the observations. This is illustrative of the IPs being able to augment the forcing-related predictability of UPs with natural-variability-related predictability by proper initialization. In these IP ensembles the transition from

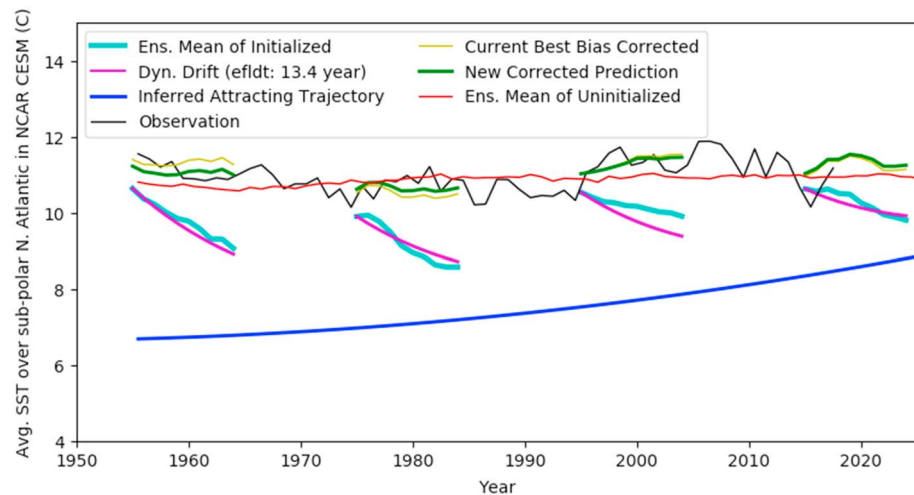


Figure 3. In this case of SST average over the subpolar North Atlantic (SPNA) in the NCAR CESM-DPLE experiment, the attracting nature of the (blue) trajectory inferred as part of the dynamic model of drift is in contrast with the initialized predictions evolving to move away from the uninitialized projection (red). This is unlike in the other cases. See Figure 2 and SI. Furthermore, in contrast to the other QoIs considered, the time scale at which this QoI loses memory is much longer. Thus, the proposed drift model raises questions about how the attracting trajectory and uninitialized projections may be reconciled, as to whether the system has multiple attractors and as to what aspects of the initialization lead to such behavior. SST = surface temperature; NCAR = National Center for Atmospheric Research; CESM = Community Earth System Model; DPLE = decadal prediction large ensemble; QoI = quantity of interest.

natural-variability-related predictability to external-forcing-related predictability is seen to occur between a year and 2 years.

The difference in behavior of the two (blue and cyan) IP ensembles is because of the different rates at which they lose memory of ICs. That is, the cyan IP ensemble is seen to remain closer to observations for a shorter period of time than the blue IP ensemble because the cyan system loses memory of the ICs faster. In this case, it may also be noted parenthetically that while the tendency of the climate system to lose memory of its initial condition and transition from natural-variability-related predictability to external-forcing-related predictability is verified in models, unlike in the weather context, the relevant time scale on which this occurs in the climate context is model dependent.

Next, in the perfect model scenario (top panel), because the model represents the various processes and small differences between such processes that lead to the observed state and its variability *perfectly*, there is no systematic deviation between the modeled state and its variability and the observed state and its variability. However, because it is difficult to achieve such levels of modeling accuracy, it is almost always the case that models exhibit biases. Because the observed mean state typically results from a dynamical balance between various processes (small differences between large numbers), errors in modeling lead to a biased (“cooler”) mean state in the lower schematic. Finally, a faster loss of memory of ICs in the cyan IP ensemble in conjunction with model bias is seen to result in an almost-jump-like behavior wherein the cyan IP trajectories are seen to quickly move away from the observations.

The main question we address in this article can also be illustrated in the schematics of Figure 1: Given the IPs in cyan, how can they be postprocessed to be more predictive of the observations in black? Answering this question leads to being able to predict near-term changes in climate better. Development of a better postprocessing method has a greater impact in the imperfect model scenario (bottom panel); CMIP models fall in this category. Our answer is based on first developing a dynamical model that describes the dominant behavior of the IPs in cyan. After validating the model, we show that the new postprocessing method better predicts observations over the hindcast period, when compared to state-of-the-art methods for postprocessing. The drift model developed suggests that a faster loss of memory of ICs, as in the cyan IPs, is more typical (cf. IPs in blue) and provides other insights that we will discuss in the final section.

2. A Dynamic Model of Drift

In the postprocessing framework, state-of-the-art methods of bias correction are purely statistical in nature and comprise mean bias correction, trend-based bias correction (TBC), and initial-state-based bias correction (e.g., Choudhury et al., 2017; Doblas-Reyes et al., 2013; Kharin et al., 2012, and others). Symbolically, given $P_{j\tau}$, a generic prediction with a start year j and at lead time τ , the corrected prediction $P_{j\tau}^{(C)}$ is obtained as

$$P_{j\tau}^{(C)} = P_{j\tau} - B_{j\tau}(P_{j\tau}, V_{j\tau})$$

where $B_{j\tau}(P_{j\tau}, V_{j\tau})$ is a statistical approximation of the bias of $P_{j\tau}$ with respect to $V_{j\tau}$ which is the verification or reference product; it is one of $B_{j\tau}^{(M)}$, or $B_{j\tau}^{(T)}$, or $B_{j\tau}^{(I)}$ for the three schemes mentioned above. Details of these techniques are given and discussed in the supporting information (SI)

In keeping with the usage of the term *initialization drift* as discussed in section 1, to be clear, when current methods of bias correction are applied to IPs, they indeed correct for such drift. What we are interested in examining is whether the dynamical aspects of initialization drift have a further useful role to play: For example, such an approach may lead to insights into IPs and eventually lead to their improvement. Or better yet, such an approach will, right away, also lead to further enhancement of the skill of IPs beyond that achieved by current bias correction techniques alone.

We now transition from the schematics of Figure 1 to considering a particular instance of near-term prediction. Here, the quantity of interest (QoI) is the average of the sea surface temperature (SST) between 60°N and 60°S in the decadal prediction large ensemble (DPLE) experiment at the National Center for Atmospheric Research (NCAR) that uses the Community Earth System Model (CESM). Not only does the climate model used in this study have a large user base, but also these recent and sophisticated experiments represent the state-of-the-art in IPs and have been made available as a community resource. (See Yeager et al., 2018, and SI for details.) In Figure 2, the cyan lines correspond to the ensemble mean of the IPs started at four different times and which extend over lead times up to 10 years. We consider a set of predictions started at 62 successive years (1955–2016), but just four of them are shown for illustration without clutter.

In Figure 2, the (rather rapid) drift of the IP trajectories away from their respective initialized states and toward a seemingly different state is evident (cf. cyan line in bottom panel of Figure 1). Next, toward developing a model of this behavior, we decompose the IP with a start year j and at lead time τ ($I_{j\tau}$) into an initialization drift $D_{j\tau}$ and a residual $\hat{I}_{j\tau}$:

$$I_{j\tau} = D_{j\tau} + \hat{I}_{j\tau} \quad (1)$$

and focus on a model for D . The model for D is derived in SI starting from a suitably averaged version of a scalar transport equation. The model for D that results from that derivation,

$$D_{j\tau} = A_{j\tau} + (D_{j0} - A_{j0}) \exp(-\alpha_j \tau), \quad (2)$$

with $D_{j0} = I_{j0}$ and $\alpha_j = \alpha_0 + \alpha_1 j$, however, can be intuitively seen to capture the behavior of IPs either in the schematic of Figure 1 or in the more realistic scenario of Figure 2: It captures the transition from natural-variability-related predictability achieved through initialization to external-forcing-related predictability in terms of losing memory of ICs at a rate of α_j per year and eventually being attracted to $A(t)$.

Given the ensemble mean of the IPs $I_{j\tau}$, the attracting trajectory $A(t)$ and parameters α_0 and α_1 are found using nonlinear least squares that minimizes the squared sum of the residual $D_{j\tau} - I_{j\tau}$ across all j s and all τ s. We consider two alternatives for inferring A : $A \equiv a_0 + a_1 j + a_2 j^2$ (Drift Model DM) or $A \equiv \{a_j\}$ (alternative DM2). While there is no conceptual difference between these two models, the statistical parsimony of DM makes it better suited when IPs are performed for only a few start years. Thus, at this point, using only the IPs $I_{j\tau}$, we have identified dynamical features that consist of (i) an attracting trajectory $A(t)$ (blue line in Figure 2 and which roughly follows the UP), (ii) a drifting component of the IP $D_{j\tau}$ that takes the initialized state at the time of initialization toward the attracting trajectory at later times (magenta lines), and (iii) a residual component $\hat{I}_{j\tau}$ (difference between cyan and magenta lines).

The drifting of IPs away from the initialized state toward the attracting state $A(t)$ conforms with our notion of the climate system forgetting its IC with time and reverting to a preferred state (climatology in the real sys-

tem; a biased state in the imperfect modeling context). At this stage, notwithstanding the fact that the drift behavior that we have successfully modeled is valid dynamical behavior, it may be worthwhile to momentarily adopt a point of view that the drift component has no further useful predictive information in it. However, on doing so, and comparing the skill of the residual component $\hat{I}_{j\tau}$ alone against traditional bias correction techniques, the poorer skill of $\hat{I}_{j\tau}$ reiterates the likelihood of the drift component containing, at least, *some* predictive information. That is, drift leads to IPs losing skill, but because it is explainable and expected dynamical behavior, it likely contains some information that can be used to enhance the skill of IPs. In the next section, we focus attention on how the validated drift model can be used to recover such information. Finally, we also note that if the behavior of the near-global mean of SST in CESM-DPLE is in any way generic, it appears that the drift away from the initialized state is rather quick—quicker than the multiyear to decadal scale at which IC-related predictability is generally thought of as transitioning over to forcing-related predictability (e.g., see Branstator & Teng, 2010, 2012; Meehl et al., 2014).

3. The Recalibrated Drift Model

The validation of the drift model in Figure 2 highlights and reiterates the point of view that one of the main problems underlying IPs is that the *model attractor* is significantly removed from the initialized state (which itself is related to observations). Consequently, one can imagine that if the (inferred) attractor is empirically corrected to be closer to the initialized state, the skill of the IPs will improve. We therefore consider a correction of the attractor $\hat{A}_{j\tau}$ given by

$$\hat{A}_{j\tau} = A_{j\tau} - B_{j\tau}^{(T)}(A_{j\tau}, I_{j\tau}^{(0)}). \quad (3)$$

Here, the corrected attractor is obtained using the trend-based bias correction with respect to predictions based solely on the IC of the IPs $I_{j\tau}^{(0)}$.

Next, since the premise of the work is that the drift of IPs is valid dynamical behavior (even though it leads to loss of predictive skill in IPs), and since we have modified the “model attractor,” it follows that the drift should be modified as well:

$$\hat{D}_{j\tau} = \hat{A}_{j\tau} + (D_{j0} - \hat{A}_{j0}) \exp(-\beta_j \tau); \quad \beta_j = \beta_0 + \beta_1 j. \quad (4)$$

We will get to the actual determination of the parameters involved momentarily, but note that, for example, β_0 and β_1 are statistical degrees of freedom. Note that this step may be justified given the observation that while the tendency of the climate system to lose memory of its initial condition is verified in models, unlike in the weather context, the relevant time scale on which this occurs in the climate context is model dependent. Finally, on having reencoded the useful component of drift $D_{j\tau}$ in the modified drift $\hat{D}_{j\tau}$, we propose that the modified drift be recombined with the residual trajectory $\hat{I}_{j\tau}$ to form the recombined (or recalibrated; see next) IP $I_{j\tau}^{(R)}$ as

$$I_{j\tau}^{(R)} = \gamma_0 \hat{I}_{j\tau} + \gamma_1 \hat{D}_{j\tau} \quad (5)$$

where γ_0 and γ_1 are further statistical degrees of freedom.

Up to this point, the only inputs to the analyses are the IPs $I_{j\tau}$. If the statistical degrees of freedom introduced in (4) and (5) are determined so as to minimize the difference between $I_{j\tau}^{(R)}$ and $I_{j\tau}^{(0)}$ in the least squares sense, it may be anticipated that $I_{j\tau}^{(R)}$ will outperform $I_{j\tau}$ in terms of skill metrics. Conventional bias correction cannot be applied at this point since observations have not been brought into the picture yet. Pragmatically, however, one is most interested, instead, in the final improvements over the hindcast period given observations. Toward examining such metrics and making comparisons to the state of the art, given the verification product $V(t)$ over the hindcast period, we correct the recalibrated prediction $I_{j\tau}^{(R)}$ using trend-based bias correction:

$$I_{j\tau}^{(RC)} = I_{j\tau}^{(R)} - B_{j\tau}^{(T)}(I_{j\tau}^{(R)}, V). \quad (6)$$

Table 1

The Use of the New Drift Correction Technique in Conjunction With Existing Bias Correction Methods Leads to Improvement in Predictive Skill of IPs, as Compared to When the IPs Are Solely Bias Corrected

QoI	Metric	Initialized decadal prediction experiment				
		TBC	RDM+TBC	%Imprv.	RDM2+TBC	%Imprv.
NCAR SST	RMSE	0.087	0.078	10	0.074	15
	ACC	0.819	0.853	4	0.868	6
NCAR ALFWC	RMSE	0.137	0.118	14	0.123	10
	ACC	0.411	0.580	41	0.550	34
NCAR SPNA	RMSE	0.357	0.327	8	0.320	10
	ACC	0.636	0.661	4	0.685	8
GFDL SST	RMSE	0.129	0.078	40	0.082	36
	ACC	0.310	0.717	131	0.679	123

Note. Both RMSE and ACC are improved (RMSE is reduced, while ACC is augmented). This is seen in each of the four cases considered with two different models (NCAR-CESM-DPLE, GFDL-CM2.1), with a surface-related QoI versus a more depth-averaged QoI (NCAR SST, NCAR ALFWC), and with a more global QoI versus a more regional QoI (NCAR SST, NCAR SPNA)—QoIs with different inertia, bias, and interannual variability characteristics. Improvements with respect to the other two commonly used bias correction techniques are given in the table in supporting information. IP = initialized predictions; TBC = trend-based bias correction; RDM = Recalibrated Drift Model; RMSE = root-mean-square error; ACC = anomaly correlation coefficient; NCAR = National Center for Atmospheric Research; SST = sea surface temperature; ALFWC = Arctic liquid freshwater content; SPNA = subpolar North Atlantic; GFDL = Geophysical Fluid Dynamics Laboratory; QoI = quantity of interest.

This is necessary, not only because there are (typically) differences between the ICs of IPs I_{j0} and observations $V(t)$ but also because we have focused on correcting IPs with respect to one dynamical aspect (drift) when there are likely other reasons for the IPs not tracking observations.

Thus, the methodology is given by (1)–(6) and involves two nonlinear least squares problems—one to find α_0 , α_1 , and A in (2) and another to find γ_0 , γ_1 , β_0 , β_1 , and parameters $b_{0\tau}$, $b_{1\tau}$ associated with TBC. In the second instance, the above mentioned parameters are determined as the arguments at the minimum of the least squares problem

$$\min_{\beta_0, \beta_1, \gamma_0, \gamma_1, b_{0\tau}, b_{1\tau}} \left\| b_{0\tau} + b_{1\tau} j - \left(I_{j\tau}^{(RC)} - V_{j\tau} \right) \right\|. \quad (7)$$

Scheme Recalibrated Drift Model (RDM) uses the quadratic approximation $A \equiv a_0 + a_1 j + a_2 j^2$, while alternative scheme RDM2 uses $A \equiv \{a_j\}$.

Next, to demonstrate the new method in action, we need to choose a specific verification product. We use Hadley Center's global Sea Ice and Sea Surface Temperature data set (Rayner et al., 2003) for the SST-related QoI being discussed and others that will be considered next and Hadley Center's EN4.2.1 (Good et al., 2013) for the Arctic-related QoI (see later). These products are widely considered as good observation-based products.

In Figure 2, the trend-based bias-corrected prediction is shown in yellow and the new RDM+TBC predictions are shown in green. Because only four (of 62) predictions are shown, it is not meaningful to interpret differences between them in Figure 2. Instead, we compute root-mean-square error (RMSE) and anomaly correlation coefficient (ACC) with respect to observations and average them over all start years and lead times as simple measures of hindcast skill. This is indicated in the first two rows of Table 1; RMSE is reduced by ~ 10 – 15% , and ACC is augmented by ~ 4 – 6% . (Table S1 has further details.)

The same computations are repeated for three other QoIs, SST averaged over the subpolar North Atlantic (SPNA)—a more regional QoI, and the Arctic liquid freshwater content (ALFWC)—a depth-integrated QoI in contrast to SST, a surface-related QoI—in the NCAR CESM DPLE experiments and for the $\pm 60^\circ$ averaged SST in a GFDL CMIP5 IP experiment (Yang et al., 2013). The corresponding values are tabulated in Table 1. The new methodology is seen to simultaneously improve both RMSE and ACC (i.e., reduce RMSE and

increase ACC) for each of the QoIs considered. Improvements with respect to the other two commonly used bias correction techniques tend to be greater, and they are given in Table S1.

4. Discussion and Conclusion

Starting from the fact that near-term predictions of climate are affected by both natural variability and external forcing, we developed a model for initialization drift, an early component of evolution of IPs that leads to a quick departure from the initialized state to a preferred (biased) state. This model permitted an identification of the preferred attracting state, the drift component along with a time scale for the loss of memory of IC, and a residual component of the IP trajectory. After recovering useful predictive information from the drift component and encoding it in the form of a modified drift, we proposed the recalibrated drift model that recombines the modified drift and the residual component in statistical fashion to minimize RMSE with respect to observations. RDM was then shown to reduce RMSE and increase ACC in a few situations when compared against current methods of bias correction. The situations considered included QoIs in two different models, an SST-related QoI versus a depth-integrated QoI, and a more global QoI versus a more regional QoI. Since the different QoIs considered exhibit different characteristics such as inertia, interannual variability, and bias, we expect broader applicability of the proposed technique.

Our modeling of the initialization-related drift provides further insights into issues related to IP. For example, in Figure 2, for the SST-related QoI considered there, the attracting trajectory inferred using only the IPs approximately follows the trajectory of the UP. The similarity is even stronger when the alternative model of drift RDM2 is used (Figure S3). However, in the same model, when a different SST-related QoI was considered that was more regional, in this case averaged over the subpolar North Atlantic, the trajectory of the inferred attracting state is far removed from the trajectory of the QoI in the UP, as seen in Figure 3. Indeed, in contrast to the attracting nature of the inferred trajectory, the IPs are seen to evolve away from the UP. Notably, here the trajectory of the QoI in the UP is closer to observations than in the IPs. On the one hand, this begs questions such as (a) Is there strong external forcing in this region, and if so what is it? (b) Does external-forcing-related predictability dominate in these regions? (c) What aspects of initialization in the subpolar region contribute to IPs being dramatically bad? On the other hand, it raises questions about how the state that corresponds to the inferred attracting trajectory and that corresponding to the UP can be dynamically reconciled or if there are multiple preferred states. The proposed drift model also helps sharpen the question of why the (average) e -folding time related to the loss of memory in the SPNA context is much longer than that for the other QoIs considered and as to why the loss of memory of ICs in the latter QoIs is much faster than has been anticipated in the literature (e.g., Meehl et al., 2014, and others). Understanding such issues raised by the analysis will likely lead to a better understanding of IP systems and may help improve them.

We anticipate that further analysis of the proposed method, for example, toward better understanding details of its working, its range of applicability, its limitations, and how it can be improved, will be useful. For example, (a) our derivation of the drift model assumes an average over a semienclosed basin, and consequently we applied the method to QoIs that were averaged or integrated over large regions. From inertia considerations alone, predictions for such QoIs have a better chance of being skillful. How much and under what circumstances can this assumption be violated and the model still be useful? What improvements are to be expected for spatially distributed fields? (b) It is encouraging that RMSE was reduced in each of the cases considered, but as to why there was also a simultaneous improvement in ACC in each of the cases is not something that we can currently explain. In this context it is also interesting to note that (Table S1) unlike with the current methods of bias correction, with the new method and its alternative, low RMSE and high ACC occur concurrently.

References

- Branstator, G., & Teng, H. (2010). Two limits of initial-value decadal predictability in a CGCM. *Journal of Climate*, 23(23), 6292–6311.
- Branstator, G., & Teng, H. (2012). Potential impact of initialization on decadal predictions as assessed for CMIP5 models. *Geophysical Research Letters*, 39, GL051974. <https://doi.org/10.1029/2012GL051974>
- Choudhury, D., Gupta, A. S., Sharma, A., Mehrotra, R., & Sivakumar, B. (2017). An assessment of drift correction alternatives for CMIP5 decadal predictions. *Journal of Geophysical Research: Atmospheres*, 122, 10–282. <https://doi.org/10.1002/2017JD026900>
- Doblas-Reyes, F., Andreu-Burillo, I., Chikamoto, Y., Garcia-Serrano, J., Guemas, V., Kimoto, M., et al. (2013). Initialized near-term regional climate change prediction. *Nature Communications*, 4, 1715.

Acknowledgments

This research was supported by the U.S. Department of Energy's (DOE) Office of Science (BER) program under a SciDAC project (PIs Bosler-Nadiga) and the Regional and Global Climate Modeling Program's HiLAT project. We thank the referees for their feedback. The presentation of the methodology has benefited from their feedback. We thank S. G. Yeager for providing the SPNA data, for facilitating access to the CESM-LE and CESM-DPLE data, and for providing useful feedback on an earlier version of the manuscript. Data analyzed in the article may be obtained from the CMIP5 archive (<https://esgf-node.llnl.gov/projects/cmip5/>) and online (<https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm4.CESM1-CAM5-DP.html>). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank GFDL for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

- Drews, A., Greatbatch, R. J., Ding, H., Latif, M., & Park, W. (2015). The use of a flow field correction technique for alleviating the North Atlantic cold bias with application to the Kiel Climate Model. *Ocean Dynamics*, 65(8), 1079–1093.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958. LLNL-JRNL-736881.
- Good, S. A., Martin, M. J., & Rayner, N. A. (2013). En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118, 6704–6716. <https://doi.org/10.1002/2013JC009067>
- Hawkins, E., Dong, B., Robson, J., Sutton, R., & Smith, D. (2014). The interpretation and use of biases in decadal climate predictions. *Journal of Climate*, 27(8), 2931–2947.
- Hazeleger, W., Guemas, V., Wouters, B., Corti, S., Andreu-Burillo, I., Doblas-Reyes, F., et al. (2013). Multiyear climate predictions using two initialization strategies. *Geophysical Research Letters*, 40, 1794–1798. <https://doi.org/10.1002/grl.50355>
- Hurrell, J., Meehl, G. A., Bader, D., Delworth, T. L., Kirtman, B., & Wielicki, B. (2009). A unified modeling approach to climate system prediction. *Bulletin of the American Meteorological Society*, 90(12), 1819–1832.
- Keenlyside, N., Latif, M., Jungclauss, J., Kornbluh, L., & Roeckner, E. (2008). Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, 453(7191), 84.
- Kharin, V., Boer, G., Merryfield, W., Scinocca, J., & Lee, W.-S. (2012). Statistical adjustment of decadal predictions in a changing climate. *Geophysical Research Letters*, 39, GL05267. <https://doi.org/10.1029/2012GL052647>
- Kim, H.-M., Webster, P. J., & Curry, J. A. (2012). Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. *Geophysical Research Letters*, 39, L10701. <https://doi.org/10.1029/2012GL051644>
- Lorenz, E. N. (1975). Climate predictability, *The physical basis of climate and climate modelling* (vol. 16, pp. 132–136), Garp publication series: WMO.
- MacLeod, D., Caminade, C., & Morse, A. (2012). Useful decadal climate prediction at regional scales? A look at the ENSEMBLES stream 2 decadal hindcasts. *Environmental Research Letters*, 7(4), 044012.
- Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, 41(9–10), 2393–2409.
- Matei, D., Pohlmann, H., Jungclauss, J., Müller, W., Haak, H., & Marotzke, J. (2012). Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *Journal of Climate*, 25(24), 8502–8523.
- 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.
- 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.
- Nadiga, B. T., Casper, W. R., & Jones, P. W. (2013). Ensemble-based global ocean data assimilation. *Ocean Modelling*, 72, 210–230. <https://doi.org/10.1016/j.ocemod.2013.09.002>
- Pohlmann, H., Jungclauss, J. H., Köhl, A., Stammer, D., & Marotzke, J. (2009). Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic. *Journal of Climate*, 22(14), 3926–3938.
- Polkova, I., Brune, S., Kadow, C., Romanova, V., Gollan, G., Baehr, J., et al. (2019). Initialization and ensemble generation for decadal climate predictions: A comparison of different methods. *Journal of Advances in Modeling Earth Systems*, 11, 149–172. <https://doi.org/10.1029/2018MS001439>
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14), 4407. <https://doi.org/10.1029/2002JD002670>
- Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., & Murphy, J. M. (2007). Improved surface temperature prediction for the coming decade from a global climate model. *Science*, 317(5839), 796–799.
- Smith, D. M., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, 41(11–12), 3325–3338.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498.
- Thoma, M., Greatbatch, R. J., Kadow, C., & Gerdes, R. (2015). Decadal hindcasts initialized using observed surface wind stress: Evaluation and prediction out to 2024. *Geophysical Research Letters*, 42, 6454–6461. <https://doi.org/10.1002/2015GL064833>
- Yang, X., Rosati, A., Zhang, S., Delworth, T. L., Gudgel, R. G., Zhang, R., et al. (2013). A predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system. *Journal of Climate*, 26(2), 650–661.
- Yeager, S., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., et al. (2018). Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bulletin of the American Meteorological Society*, 99(9), 1867–1886.