Decadal climate predictions with the Canadian Earth System Model version 5 (CanESM5)

Reinel Sospedra-Alfonso, William J. Merryfield, George J. Boer, Viatsheslav V. Kharin, Woo-Sung Lee, Christian Seiler, and James R. Christian

Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, University of Victoria, Victoria, BC, V8N 1V8

Correspondence: Reinel Sospedra-Alfonso (reinel.sospedra-alfonso@canada.ca)

Abstract.

The Canadian Earth System Model version 5 (CanESM5) developed at Environment and Climate Change Canada's Canadian Centre for Climate Modelling and Analysis (CCCma) is participating in phase 6 of the Coupled Model Intercomparison Project (CMIP6). A 40-member ensemble of CanESM5 ~~historical~~ retrospective decadal forecasts (or hindcasts) is integrated for ten years from realistic initial states once a year during 1961 to present using prescribed external forcing. The results are part of CCCma's contribution to the Decadal Climate Prediction Project (DCPP) component of CMIP6. This paper evaluates CanESM5 large ensemble decadal hindcasts against observational benchmarks and against historical climate simulations initialized from pre-industrial control run states. The focus is on the evaluation of the potential predictability and actual skill of annual and multi-year averages of key oceanic and atmospheric fields at regional and global scales. The impact of initialization on prediction skill is quantified from the hindcasts decomposition into uninitialized and initialized components. The dependence of potential and actual skill on ensemble size is examined. CanESM5 decadal hindcasts skilfully predict upper-ocean states and surface climate with a significant impact from initialization that depend on climate variable, forecast range, and geographic location. Deficiencies in the skill of North Atlantic surface climate are identified and potential causes discussed. The inclusion of biogeochemical modules in CanESM5 enables the prediction of carbon cycle variables which are shown to be potentially skilful on decadal time scales, with a strong long-lasting impact from initialization on skill in the ocean and a moderate short-lived impact on land.

1 Introduction

The Canadian Earth System Model version 5 (CanESM5) is the latest Canadian Centre for Climate Modelling and Analysis (CCCma) global climate model. ~~It~~ CanESM5 has the capability to incorporate~~s~~ an interactive carbon cycle, and was developed to simulate historical climate change and variability, to make centennial scale projections of future climate, and to produce initialized climate predictions on seasonal to decadal time scales (Swart et al., 2019b, this issue, hereafter S19). S19 summarizes CanESM5 components and their coupling, together with the model's ability to reproduce large scale features of the historical climate and its response to external forcing. This paper examines the predictive ability of CanESM5 on decadal time scales.

CanESM5 decadal climate predictions are CCCma contribution to Component A of the Decadal Climate Prediction Project (DCPP; Boer et al., 2016) endorsed by phase 6 of the Coupled Model Intercomparison Project (CMIP6). ~~In this paper, the skill of key physical climate and carbon cycle variables is assessed.~~

The aim of ~~near-term or~~ decadal climate predictions is to provide end users with useful climate information, on time scales ranging from one year to a decade, that improves upon the information obtained from climate simulations that are not initialized from observation-based states (Merryfield et al., 2020). On decadal timescales, the evolution of the climate depends on the interplay between an externally forced component (e.g., resulting from the changes in greenhouse gas emissions, aerosols, land use, solar forcing) and internally generated natural variability. The climate response to the externally forced component may be estimated from climate simulations that are not initialized from observation-based states (henceforth referred to as "uninitialized simulations"), but the internally generated component of these simulations is not expected to match observations. For decadal predictions, climate models are initialized from observation-based states and integrated forward for several ~~up to 10~~ years with prescribed external forcing. The expectation is that by taking advantage of predictable slowly-varying internally generated fluctuations of the climate system, including those originating from the ocean, climate phenomena such as multi-year atmospheric circulation changes (Smith et al., 2010; Sutton and Dong, 2012; Monerie et al., 2018), their impact on near-surface temperature and rainfall (Zhang and Delworth, 2006; Boer et al., 2013; McKinnon et al., 2016; Sheen et al., 2017), and the frequency of extreme weather events (Eade et al., 2012; Ruprich-Robert et al., 2018), can be predicted a year or more in advance. Furthermore, initialization affects the model response to external forcing and can potentially correct unrealistic simulated trends (Sospedra-Alfonso and Boer, 2020).

For decadal predictions to be credible they must be accompanied by measures of historical skill, and an understanding of the various processes that contribute to skill on decadal time scales (Meehl et al., 2009). CMIP5 (Taylor et al., 2012) provide a framework for quantifying the "added value" of initialized climate predictions over simulations (Meehl et al., 2009; Goddard et al., 2013) and, building upon this experience, the DCPP panel coordinated a comprehensive set of decadal prediction experiments endorsed by CMIP6. These include ensembles of historical initialized predictions (hindcasts) and historical simulations as well as basic guidelines for post-processing model output (Boer et al., 2016). Assessing the added value of decadal climate predictions over climate simulations can, however, be difficult. In the presence of a strong long-term climate response to external forcing, as is the case for near-surface air temperature for instance, the externally forced component carries most of the predictable variance and can mask the contribution of relatively weaker internal variations to the skill (Smith et al., 2019; Sospedra-Alfonso and Boer, 2020). Underestimation of the predictable signal can also degrade decadal forecast skill (Sienz et al., 2016), motivating the use of large ensembles to better extract the predictable component of the forecast (Scaife and Smith, 2018; Yeager et al., 2018; Smith et al., 2019, 2020; Deser et al., 2020).

In this work, we assess the potential predictability and actual skill of 40-member ensemble hindcasts produced using CanESM5. The hindcasts are initialized ~~at the end~~ on December 31st of every year during 1960 to 2019 and run for 10 years. The potential predictability framework enables the detection of the climate model "signal" that is common to the ensemble of model predictions whereas actual skill refers to the forecast ability to reproduce the predictable signal of the climate system. The initial ensemble spread is intended to represent observational uncertainties (Merryfield et al., 2013) and to be small compared with the amplitude of the climate signal to be predicted. Forecast skill is compared here against climatology, persistence, and ~~historical~~ uninitialized simulations. The contribution to skill of initialized and uninitialized components of the forecasts are inferred with reference to a 40-member ensemble of ~~historical~~ uninitialized simulations~~,~~ spanning 1850–2014. Both hindcasts and uninitialized simulations use the same external forcing parameters during their overlapping period. The dependence of potential and actual skill on ensemble size at regional and global scales is also examined.

This paper serves to document the CanESM5 decadal hindcasts that are the CCCma contribution to Component A of the DCPP (Boer et al., 2016). It highlights CCCma's newly developed capabilities including prediction of biogeochemical ~~variables and the~~ and carbon cycle variables, as well as the use of large ensembles to better extract the predictable component of the forecasts. These are steps towards a more comprehensive decadal climate prediction system at CCCma, although not without new challenges and deficiencies, some of which are examined here. The remainder of the paper provides an overview of CanESM5 model components (section 2), a description of the initialization process and ensemble generation (section 3), the methodology for hindcasts evaluations (section 4) together with supporting information (appendices A and B), and assessments of potential predictability and actual skill in the upper ocean (section 5) and the surface climate on land (section 7). Section 6 addresses issues of the sea surface temperature hindcasts in the subpolar North Atlantic and Labrador Sea. The impact of ensemble size on potential predictability and actual skill is discussed in section 8, whereas the ~~predictability~~ potential for skilful predictions of carbon cycle variables is examined in section 9. Section 10 provides a summary and the conclusions. CanESM5 output data used here including hindcasts, assimilation runs to initialize hindcasts, volcanic experiments, and historical uninitialized simulations are freely available from the Earth System Grid Federation on https://esgf-node.llnl.gov/search/cmip6/.

2 The CanESM5 eEarth system model

A detailed description of CanESM5 and its components is given in S19 and the references therein, thus we provide only a brief summary here. CanESM5 couples version 5 of the Canadian Atmosphere Model (CanAM5; Cole and Coauthors, 2019, this issue) with the CanNEMO ocean component adapted from the Nucleus for European Modelling of the Ocean version 3.4.1 (NEMO3.4.1 Madec and Coauthors, 2012). CanAM5 incorporates the Canadian Land Surface Scheme version 3.6.2 (CLASS3.6.2; Verseghy, 2000) and the Canadian Terrestrial Ecosystem Model (CTEM; Arora and Boer, 2005) whereas CanNEMO represents ocean biogeochemistry with the Canadian Model of Ocean Carbon (CMOC; Zahariev et al., 2008; Christian et al., 2010). Sea ice is simulated within the NEMO framework using version 2 of the Louvain-la-Neuve Sea Ice Model (LIM2; Fichefet and Maqueda, 1997; Bouillon et al., 2009). CanAM5 and CanNEMO are coupled with CanCPL developed at CCCma to facilitate communication between the two components.

CanAM5 is a spectral model with a T63 triangular truncation and 49 hybrid vertical coordinate levels extending from the surface to 1hPa. Physical quantities are computed on the linear transform grid leading to a horizontal resolution of approximately 2.8 degrees. Improvements of CanAM5 upon its predecessor CanAM4 (von Salzen et al., 2013) include the addition of 14 vertical levels in the upper troposphere and stratosphere, upgraded treatment of radiative processes, particularly in the parameterization of albedo for bare soil, snow and ocean white-caps, improved aerosol optical properties, better optical properties for ice clouds and polluted liquid clouds, and a more comprehensive representation of land surface and lake processes. CanNEMO is configured on the ORCA1 C-grid with 45 vertical levels with vertical spacing ranging from about 6 meters near the surface to about 250 meters in the abyssal ocean. The horizontal resolution is based on a 1 degree isotropic Mercator grid which is refined meridionally to 1/3 of a degree near the Equator, and includes a tripolar configuration to avoid the coordinate singularity in the Northern Hemisphere.

CanESM5 ~~has the capability for an interactive carbon cycle and thus~~ includes biogeochemistry modules to simulate land and ocean carbon exchange with the atmosphere. For the land surface, CLASS simulates energy, water, and momentum fluxes at the land-atmosphere boundary whereas CTEM simulates atmosphere-land fluxes of CO_2 and related terrestrial processes including photosynthesis, autotrophic and heterotrophic respiration, leaf phenology, carbon allocation, biomass turnover and conversion of biomass to structural attributes (Arora, 2003; Arora and Boer, 2003, 2005). This enables CTEM to simulate gross and net primary productivity over land while tracking the carbon flow through three living vegetation components (leaves, stem and roots) for nine plant functional types of prescribed fractional coverage (Melton and Arora, 2016), and two dead carbon pools (litter and soil). For the ocean, CMOC simulates carbon chemistry and abiotic chemical processes (such as solubility of oxygen, inorganic carbon, nutrients and other passive tracers having no feedback on biology and the simulated climate) in accordance with the CMIP6 Ocean Model Intercomparison Project (OMIP) biogeochemical protocol (OMIP-BGC; Orr et al., 2017). The biological module of CMOC is a simple Nutrient-Phytoplankton-Zooplanston-Detritus model, with fixed Redfield stoichiometry, and simple parameterizations of iron limitation, nitrogen fixation, and export flux of calcium carbonate. Both ~~initialized and uninitialized predictions~~ hindcasts and uninitialized simulations examined here, however, have prescribed atmospheric CO_2 concentrations and thus ocean and land CO_2, being purely diagnostic, do not feed back onto the simulated physical climate.

3 Forcing, initialization and ensemble generation

External forcing agents including historical anthropogenic and natural greenhouse gases, volcanic aerosols, solar activity and land use change are specified according to the CMIP6 protocol (Eyring et al., 2016). Emissions of sulfur dioxide (SO2), dimethyl sulfide (DMS), black carbon, and organic carbon aerosol are specified, whereas mineral dust and sea salt emissions are simulated depending on local conditions. Concentrations of oxidants are specified for simulations of oxidation of sulfur species in clear air and in clouds. Direct effects of all types of aerosols, and 1st and 2nd indirect effects of sulfate, are simulated. Beyond the historical period, forcing from the Shared Socioeconomic Pathway (SSP) 2-4.5 scenario is used.

~~For each ensemble member, the coupled model is initialized from a separate assimilation run that ingests observation-based data from the ocean, atmosphere and sea ice as detailed below. These data-constrained assimilation runs span 1958 to present and are started from consecutive years following a 80-year spinup run that assimilates repeating 1958–1967 data. This leads to a spread of initial states that represent observational uncertainties. Full-field initialization is employed, and a standard lead time-dependent bias correction is applied in calculating forecast anomalies (Boer et al., 2016).~~ A full-field initialization method is used to provide the initial conditions of the hindcasts. Each hindcast member is initialized from a separate assimilation run that ingests observation-based data from the ocean, atmosphere and sea ice in a coupled-model mode as detailed below. Each data-constrained assimilation run branches off a long spinup run used to quasi-equilibriate the physical and biogeochemical model states by assimilating repeating 1958-1967 data. After an 80-year spinup, one assimilation run is started every year for 40 more spinup years to produce a 40-member ensemble of assimilation runs that are run from 1958 until present. The differences in assimilation run initial conditions combined with the insertion of only $1/4$ of the atmospheric analysis increment as described below, lead to assimilation runs that are not identical despite of assimilating the same observation-based data. This leads to a spread of initial states for the hindcasts that represent observational uncertainties.

For the global ocean, 3D potential temperature and salinity are nudged toward values interpolated from monthly Ocean Reanalysis System 5 (ORAS5; Zuo et al., 2019) with a 10 day time constant in the upper 800m, and 1 year time constant at greater depths. The 1°S-1°N band is excluded partly to avoid disturbing strong equatorial currents below the surface (Carrassi et al., 2016). Sea surface temperature is relaxed to daily values interpolated from weekly values of the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation Sea Surface Temperature (OISST; Banzon et al., 2016) during November 1981 to present, or monthly values from the NOAA's Extended Reconstructed Sea Surface Temperature (ERSSTv3; Xue et al., 2003; Smith et al., 2008) during 1958 to October 1981, with a 3 day time constant.

Sea ice concentration is relaxed to daily values interpolated from monthly values of the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST.2; Titchner and Rayner, 2014) ~~augmented with~~ merged with weekly data from digitized Canadian Ice Service charts from 1958 to 2014 (Tivy et al., 2011), and to daily values from the Canadian Meteorological Centre (CMC) analysis from 2015 to present, with a 3 day time constant. Sea ice thickness is relaxed to daily-interpolated monthly values from the SMv3 statistical model of Dirkson et al. (2017) with a 3 day time constant. Before 1981, a repeating 1979–1988 monthly climatology is used.

Atmospheric temperature, horizontal wind components and specific humidity are nudged toward the European Centre for Medium-Range Forecasts (ECMWF) 6-hourly ERA-Interim (Dee et al., 2011) reanalysis values during 1979-present, or ERA40 (Uppala et al., 2005) anomalies added to ERA-Interim climatology during 1958-1978. The relaxation of the atmospheric variables is done with a 24 hour time constant (corresponding to inserting only $1/4$ of analysis increment) and excludes spatial scales smaller than about 1000km. This results in ensemble spread that is comparable to root mean square differences between different atmospheric reanalyses (Merryfield et al., 2013).

~~Land physical and biogeochemical (BGC) variables are initialized through response of CLASS-CTEM to the data-constrained atmosphere.~~ The land physical and biogeochemical (BGC) variables in the assimilation runs, which provide the initial values for the land variables in the hindcasts, are not directly constrained to observations but are determined by the CLASS-CTEM response to the evolving data-constrained atmosphere of the coupled model. Land carbon pools are spun up during the 80-year spinup mentioned above. Similarly, ~~O~~oceanic BGC variables are initialized through response of CMOC to evolving data-constrained physical ocean variables and surface atmospheric forcing.

To briefly assess the effect of natural aerosols on precipitation, we employ the volcanic experiments prepared for contribution to Component C3 of the DCPP (Boer et al., 2016). These are two sets of three separate experiments, with and without volcanic forcing. The experiments without volcanic forcing repeat the 1963, 1982 and 1991 hindcasts, except that the stratospheric aerosols are specified as per the 2015 hindcasts. These experiments exclude the effects of volcanic aerosols on the simulated climate due to the Agung, El Chichón and Pinatubo events of those years, respectively. The three experiments with volcanic forcing repeat the 2015 hindcasts with stratospheric aerosols from the 10-year period starting in 1963, 1982 and 1991, respectively, and represent the impact of these volcanic events on different climate states.

4 Hindcasts evaluation methods

The evaluation approach and notation largely follows Boer et al. (2013, 2019a, b) and Sospedra-Alfonso and Boer (2020, hereafter SB20) where observations X, ensemble of ~~forecasts~~ hindcasts Y_k, and ensemble of uninitialized simulations U_k are annual or multi-year mean anomalies that are functions of time and location. The sub-index $k = 1 \dots m_Y$ or $k = 1 \dots m_U$ denotes ensemble member, where m_Y and m_U represent the ensemble size of ~~forecasts~~ hindcasts and uninitialized simulations, respectively. The anomalies are computed relative to climatological averages over a specified time period that is common to model output and observations. For the ~~forecasts~~ hindcasts and uninitialized simulation ensembles, the anomalies are represented as

$$
\begin{aligned}
Y_k &= \Psi + y_k = \psi_f + \psi + y_k \\
U_k &= \Phi + u_k = \phi_f + u_k
\end{aligned}
\tag{1}
$$

consisting of predictable or "signal" components (Ψ, Φ) and unpredictable or "noise" components (y_k, u_k). The predictable components are, in turn, comprised of externally forced (ψ_f, ϕ_f) and internally generated ψ variability. Even though the ~~forecasts~~ hindcasts and uninitialized simulations see the same external forcing, their forced components are not generally the same because initialization affects both the forced response and the internally generated variability. Unlike the ~~forecasts~~ hindcasts, internal variability in the uninitialized simulations is not constrained by initialization and is not predictable. The predictable components are common across the ensemble while the unpredictable components (y_k, u_k) differ and average to zero over a large enough ensemble. The assumption is that all variables average to zero over the time period considered, forced components are independent of internally generated components, and all are independent of the noise components.

Ensemble averaging across ~~forecasts~~ hindcasts and uninitialized simulations in Eq. (1), denoted here by dropping the sub-index k, leads to the following representation of the ensemble mean ~~forecasts~~ hindcast and ensemble mean simulation

$$
\begin{aligned}
Y &= \Psi + y = \psi_f + \psi + y \longrightarrow \psi_f + \psi \\
U &= \Phi + u = \phi_f + u \longrightarrow \phi_f
\end{aligned}
\tag{2}
$$

where here and elsewhere the arrows indicate the large ensemble limit. ~~The variances of the ensembles of forecasts and simulations in Eq. (1) are, respectively, $\sigma^2_{Y_e} = \sigma^2_\Psi + \sigma^2_{y_e}$ and $\sigma^2_{U_e} = \sigma^2_\Phi + \sigma^2_{u_e}$, while that of the ensemble means in Eq. (2) are $\sigma^2_Y = \sigma^2_\Psi + \sigma^2_{y_e}/m_Y$ and $\sigma^2_U = \sigma^2_\Phi + \sigma^2_{u_e}/m_U$, where $\sigma^2_{y_e}$ and $\sigma^2_{u_e}$ are the noise variances of forecasts and simulations. These variances are given explicitly in Eqs. (A5)-(A10) in appendix A.~~ The total variances $\sigma^2_{Y_e}$ and $\sigma^2_{U_e}$ of the ensembles of forecasts and simulations in Eq. (1), denoted with subscript e for "ensemble", and the variances σ^2_Y and σ^2_U of the ensemble means in Eq. (2), are given explicitly in Eqs. (A1)-(A10) of appendix A.

SB20 decompose Ψ into mutually independent uninitialized Y_u and initialized Y_i components, with

$$Y = Y_u + Y_i + y \tag{3}$$

where the component $Y_u = \alpha\phi_f$ is ascribed to uninitialized external forcing ~~(here α is a measure of the projection of ψ_f on ϕ_f),~~ while the initialized component $Y_i = (\psi_f - \alpha\phi_f) + \psi$ includes the effect of initialization on both the forced component and the unforced internally generated component. Here α is the regression coefficient of ψ_f and ϕ_f, which is set to zero when the covariance of ψ_f and ϕ_f is negative (Eq. 18; SB20). The potentially predictable variance fraction (*ppvf*, Boer et al., 2013, 2019a, b) of the ~~forecast~~ hindcast ensemble and that of the ensemble mean ~~forecast~~ hindcast are, respectively

$$q_e = \frac{\sigma_\Psi^2}{\sigma_{Y_e}^2} = \frac{\sigma_{Y_u}^2 + \sigma_{Y_i}^2}{\sigma_{Y_e}^2} = q_{e_u} + q_{e_i} \tag{4}$$

$$q = \frac{\sigma_\Psi^2}{\sigma_Y^2} = \frac{\sigma_\Psi^2}{\sigma_\Psi^2 + \sigma_{y_e}^2/m_Y} = \frac{1}{1+\gamma_Y/m_Y} = \frac{\sigma_{Y_u}^2 + \sigma_{Y_i}^2}{\sigma_Y^2} = q_u + q_i \longrightarrow 1 \tag{5}$$

where q depends on ensemble size and both q and q_e are less than one. Here, $\gamma_Y = \sigma_{y_e}^2/\sigma_\Psi^2$ is the noise-to-predictable variance ratio of the ~~forecast~~ hindcast ensemble, Eq. (A12). The *ppvf* of the simulations is defined in like manner. The *ppvf*'s q_e and q represent, respectively, the fractions of total and ensemble mean forecast variances that are potentially predictable. "Potential predictability" refers here to predictability within the "model world", i.e., to predictability of a signal that is expected to represent variations of the observed climate system, but which may be unrealistic due to model and/or initialization errors. A potentially predictable signal is necessary but not sufficient for actual skill. The uninitialized and initialized contributions to q, denoted q_u and q_i, respectively, are computed explicitly according to Eqs. (A14)-(A15) in appendix A.

Following SB20, the correlation skill (or anomaly correlation coefficient) r_{XY} of the ensemble mean ~~forecast~~ hindcast can be decomposed as

$$r_{XY} = r_{XY_u}\frac{\sigma_{Y_u}}{\sigma_Y} + r_{XY_i}\frac{\sigma_{Y_i}}{\sigma_Y} = r_u + r_i \tag{6}$$

where r_{XY_u} and r_{XY_i} are the correlation skills of the uninitialized and initialized components Y_u and Y_i themselves, while r_u and r_i are the contributions to the overall correlation skill obtained by scaling with the fractions $\sqrt{q_u}$ and $\sqrt{q_i}$ representing the variances involved. This decomposition allows the assessment of the impact of initialization on correlation skill and explicitly accounts for the effects of initialization on the model response to external forcing, through r_i, and by excluding the comparatively strong contribution to variability by the trends, through r_{XY_i}, which can obscure predictable internal variations. The latter avoids having to linearly detrend the data, which is frequently done and can introduce errors (SB20). The explicit computation of r_{XY_u} and r_{XY_i} as well as r_u and r_i can be found in SB20 and are given in Eqs. (A16)-(A19) of the appendix for completeness.

The potential correlation skill of the ~~forecasts~~ hindcasts is (Boer et al., 2013, 2019a, b)

$$\rho = \frac{\sigma_\Psi^2}{\sigma_{Y_e}\sigma_Y} = \sqrt{\beta}q \longrightarrow \sqrt{q_e} \tag{7}$$

where $\beta = \sigma_Y^2/\sigma_{Y_e}^2 < 1$. The squared potential skill gives the fraction of the ensemble total variance that is represented or "explained" by the ensemble mean ~~forecast~~ hindcast, which in the large ensemble limit is the *ppvf* q_e. The connection between the potential and actual correlation skill has been discussed by Eade et al. (2014); Smith et al. (2019); Strommen and Palmer (2019)~~, where the following~~ in terms of a ratio of predictable components ~~is considered~~

$$\Pi_r \;=\; \frac{r_{XY}}{\sqrt{\beta}} = \frac{r_{XY}}{\rho}\frac{\rho}{\sqrt{\beta}} = \frac{r_{XY}}{\rho}q \longrightarrow \frac{r_{XY}}{\sqrt{q_e}} = r_{XY}\sqrt{1+\gamma_Y} \tag{8}$$

A similar quantity can be defined for the simulations. Assuming $r_{XY} > 0$, if $r_{XY}/\sqrt{\beta} > 1$ then $r_{XY} > \rho$ since $q < 1$ and the actual correlation skill exceeds potential skill (Boer et al., 2019b), i.e., the model is more skilful at predicting the observations than its own behaviour. In the large ensemble limit, this is possible only if the noise-to-predictable variance ratio γ_Y is large enough to offset the correlation skill of the ensemble mean prediction, and can occur when the forecast predictable variance is much smaller than the observed variance. Such a behaviour is referred to as signal-to-noise paradox by Scaife and Smith (2018).

Evaluations of hindcasts actual skill also include computations of the mean square skill score

$$\text{MSSS}(Y,R,X) = 1 - \frac{\text{MSE}(Y,X)}{\text{MSE}(R,X)} \tag{9}$$

where MSE(Y,X) and MSE(R,X) are the mean square errors of, respectively, ~~forecasts~~ hindcasts and reference predictions relative to observations (Goddard et al., 2013; Yeager et al., 2018). The reference predictions used here are the climatology of the observed anomalies $\overline{X} = 0$, persistence X_p, and the uninitialized ~~predictions~~ simulations, U. For evaluations of N-year mean ~~forecasts~~ hindcasts we use observed N-year rolling averages over the forecast initialization years. Persistence equals the most recent observed N-year average at the time of forecast initialization, and the uninitialized ~~predictions~~ simulations are the N-year rolling averages of the ensemble mean historical simulations. We evaluate N-year averages of annual or seasonal mean anomalies including $N = 1$ for the second year of the ~~forecast~~ hindcast (Year 2) and $N = 4$ for ~~forecast~~ hindcast years 2 to 5 (Year 2-5) and 6-9 (Year 6-9) corresponding to forecast ranges beyond seasonal lead times. Anomalies are taken relative to identically sampled climatologies and predictions are bias corrected (but not trend corrected) following the recommendations of Boer et al. (2016). Annual averages are taken from January to December and seasonal averages are as specified in each case. Because hindcast initialization is done in late December, winter averages (DJF) predictions of, say Year 2, correspond to December of Year 1 ~~forecasts~~ hindcasts and January and February of Year 2.

Statistical significance is evaluated using a non-parametric moving-block bootstrap approach (Goddard et al., 2013; Wilks, 1997) to generate the skill score's sampling distribution based on 1000 repetitions. For every grid cell, skill scores are generated by resampling the data, with replacement, along the time dimension, and along the ensemble members dimension for hindcasts and simulations. Following Goddard et al. (2013), resampling of 5-year blocks are considered to account for temporal autocorrelation. The 5%- and 95%-quantile estimates of the bootstrapping distribution of the skill scores determine the 90% confidence interval. If the confidence interval does not include zero, the skill score is deemed statistically significant with 90% confidence and the associated grid cell is cross-hatched in the maps. The Fisher's Z-transformation is applied to correlation skill scores before computing confidence intervals and its inverse is applied to the resulting quantiles.

5 Predictability and skill in the upper ocean

The time evolution of the upper-ocean heat content (OHC) is modulated by a wide range of low-frequency variability ranging from decadal and multidecadal to ~~secular~~ centennial or longer time scales (Levitus et al., 2005; Taguchi et al., 2017). The ocean's lagged response to atmospheric thermal and dynamical forcing is due to the high specific heat capacity of water, which makes the upper ocean a major source of surface climate predictability on seasonal to decadal time scales (e.g., Smith et al., 2007; Meehl et al., 2014; Yeager and Robson, 2017). The top panels of Fig. 1 show the *ppvf* q_e, Eq. (4), of heat content in the upper 300 m of the ocean (OHC300) for forecast years 2, 2-5 and 6-9. For Year 2, $q_e > 0.7$ in most of the global ocean, implying that the OHC300 predictable variance accounts for more than 70% of the total variance. This contrasts with sectors of the tropical Pacific, the Indo-Pacific warm pool and in some coastal regions, where it can be lower than 30% (Fig. 1a). The relatively lower predictability and small effect of initialization in equatorial regions (Fig. 1a,d) may be partly associated with the 3D ocean initialization procedure that excludes the 1°S-1°N band (section 3), and to the fast wave processes on the equator. By contrast, initialization has a strong impact on q_e in vast extratropical regions (Fig. 1d). For multi-year averages, regions of lower q_e extend to the subtropics and to higher latitudes in the North Pacific, particularly for larger lead times (Fig. 1b,c). The impact of forecast initialization is widespread for Year 2-5 (Figs. 1e), but is much reduced for Year 6-9 (Figs. 1f) when most potentially predictable variance (Figs. 1c) is attributed to the simulated external forcing. A few notable exceptions showing a persistently high initialized potentially predictable variance include the North Atlantic, the Arctic and sectors of the Southern Ocean.

Some of the predictable variance contributes to skill, but some may reflect model biases and/or initialization errors. The correlation skill r_{XY} of OHC300 hindcasts and the contribution r_i from initialization are shown in the upper and lower panels of Fig. 2, respectively. For Year 2, correlation skill is significant over large portions of the global ocean (Fig. 2a), and is reduced in some extratropical regions including sectors of the eastern Pacific, the Arctic and Southern oceans, the Alaskan and western subarctic gyres, and in sectors of the North Atlantic most notably the western sub-polar region (WSPNA) and the Labrador Sea. The negative skill in the WSPNA region is attributed to initialization (Fig. 2d) and is partly a consequence of erroneous trends in ORAS5 reanalysis (Johnson et al., 2019) being imprinted on the hindcasts. The poor skill in the WSPNA and plausible causes are discussed further in section 6 below. Positive contributions from initialization can be seen in large sectors of the Pacific and Indian ocean basins for Year 2, whereas correlation skill in the Atlantic results mostly from uninitialized external forcing (Figs. 2a,d). For multi-year averages (Fig. 2b,c), the geographic extent of positive correlation skill is somewhat reduced relative to Year 2, most notably over the equatorial Pacific and in the Indian basin at longer leads, and tends to increase in magnitude over regions where skill is attributed to the simulated external forcing.

Sea surface temperature (SST) ~~forecasts~~ hindcasts for Year 2, 2-5 and 6-9 (Figs. 3a-c) show *ppvf* $q_e > 0.4$ in most of the global ocean, with larger values for the multi-year averages. For Year 2, notable contributions from initialization $q_{e_i} > 0.3$ are seen in the western equatorial Pacific, which are not present for the multi-year averages. ~~Excluding the Arctic region, t~~The predictable SST signal is strongest in the Arctic, in sectors of the Southern Ocean, and in the WSPNA and Labrador Sea regions, resulting entirely from initialization (Fig. 3d-f). These locations are characterized by unrealistic ~~strong~~ negative trends

~~consistent with those from~~ in the ORAS5 reanalysis (Fig. 4b)~~.~~, that are imprinted in the hindcasts (Fig. 4d) and contribute to the predictable variance attributed to initialization (Fig. 3d-f). On inter-annual time scales, q_e and q_{e_i} are generally smaller for SST than OHC300 (Figs. 1a-c), as SST is more directly affected by atmospheric conditions. On multi-year time scales and longer lead times, q_e ~~is generally~~ can be larger for SST (Fig. 3c) than OHC300 (Fig. 1~~d-f~~c), as the simulated forced component becomes dominant, which strongly impacts SST trends (Fig. 4c).

SST ~~forecasts~~ hindcasts show a reasonably widespread correlation skill (Fig. 5a-c). A large fraction of SST correlation skill is attributable to the uninitialized external forcing, but significant contributions r_i from initialization are seen in all ocean basins for Year 2 (Figs. 5d), and in sectors of the Pacific and Southern Ocean for the multi-year averages (Figs. 5e-f). The large contribution to skill by the uninitialized component derives partly from temperature trends that account for a larger variance fraction than that of the initialized component (Figs. 3), which is itself skilful as per r_{XY_i} (Figs. 5g-i). This shows that despite the significant correlation skill r_{XY_i} of the initialized component Y_i, the associated small variance fraction q_{e_i} (Fig. 3) reduces the contribution $r_i = \sqrt{q_{e_i}} r_{XY_i}$ (Eqs. 4 and 6) of the initialized component to correlation skill. The apparent skill reemergence of the initialized component in the eastern Pacific for Year 6-9 is noteworthy (Fig. 5i).

MSSS of SST hindcasts relative to observed climatology (Figure 6a-c) indicate significant skill in large sectors of the North Atlantic, in the Indian Ocean, and in the western Pacific extending into the southeast extratropics. MSSS is a more stringent measure than correlation skill, and regions with significant correlation skill but near zero or negative MSSS indicate a misrepresentation of the observed variance for the given linear relationship between predictions and observations (i.e., due to a conditional bias). This is the case for various extratropical regions and sectors of the tropical Pacific (compare Figs. 5a-c and 6a-c). Regions with MSSS<<0 indicate a disproportionally large predicted variance relative to observations, such as in WSPNA and Labrador Sea, the Arctic, and in the Southern Ocean (Figs. 6a-c). These regions are characterized by high q_e (Figs. 3a-c) with a strong impact from initialization (Figs. 3d-f), but lack actual skill (Figs. 5a-c and 6a-c). For Year 2, SST hindcasts outperform persistence in most of the tropics (Fig. 6d), except for sectors of the subequatorial and western Pacific, and the western South Atlantic. For the multi-year averages (Fig. 6e,f), SST hindcasts beat persistence in large sectors of the Atlantic and Indian oceans, and in most western and southern portions of the Pacific within the 40°S-40°N latitude band. The hindcasts outperform the uninitialized simulations, particularly for multi-year averages (Figs. 6h,i), in vast subequatorial regions, in the Indian Ocean, and in northern and subpolar regions, but underperform in sectors of the Sourthern Ocean, the eastern and south Atlantic, and in the WSPNA and Labrador Sea regions.

6 Erroneous SST hindcasts in the WSPNA and Labrador sea regions

The negative correlation skill in the WSPNA and Labrador Sea for both upper-ocean heat content (Fig. 2) and SST (Fig. 5) is fully attributed to initialization (i.e., $r_{XY} = r_i$) indicating a mismatch between the external forced response in the hindcasts and uninitialized simulations ~~responses to external forcing in these regions~~ (SB20), which can be different as implied by Eq. (1). CanESM5 ocean is initialized with ORAS5 (section 3), which has unrealistic temperature and salinity trends in the upper subpolar North Atlantic associated with erroneous water mass and heat transport before the 2000s (Jackson et al., 2019;

Tietsche et al., 2020). The Labrador Sea in ORAS5 presents large changes in density anomaly, most notably in deep waters (1500-1900 m), which decrease abruptly from late 1990's to early 2000's leading also to unobserved trends (Fig. 9 of Jackson et al., 2019). These variations and unrealistic trends are imprinted on CanESM5 assimilation runs as they are nudged toward ORAS5 temperature and salinity fields to initialize the hindcasts (section 3). The anomalous heat and saline surface water transport into WSPNA is largely compensated in ORAS5 by a strong surface cooling provided by relaxation to observed SST (Tietsche et al., 2020), but such a cooling is not present in the forecasts, which leads to imbalances. This, combined with the model inherent biases in the region (S19) and resulting forecast drift, yield unrealistic decadal variations and long-term trends in the hindcasts themselves, which affect skill.

Figure 7a shows January-February-March (JFM) time series and linear trends of SST averages over WSPNA (40°N–60°N,50°W–30°W) for ERSSTv5 as verifying observations, ORAS5, and CanESM5 assimilation runs, Year 1 and Year 2 hindcasts, and the simulations. Analogous plots for averages over the Labrador Sea (55°N–65°N,60°W–45°W) are shown in Fig. 8. In WSPNA, ORAS5 is warmer than observations during the mid 1970's to about 2000's, as is the case for the assimilation runs. Year 1 hindcasts remain close to initial SST for most years although are somewhat colder during late 1990's and onward. Year 2 SST hindcasts, on the other hand, have a strong warming in the early 1970's and remain 2-3°C warmer than observations until late 1990's, when a steep cooling occurs to below observed values until early 2000's. These changes yield a negative trend for Year 2 hindcasts (-0.02 °C/decade) that does not match the slight warming trend from observations (0.01 °C/decade). By comparison, ORAS5 and the assimilation runs have virtually no trend, with values of 0.002 and 0.004 °C/decade, respectively. For longer lead times, the hindcasts drift toward simulations (not shown), which are characterized by a strong cold bias (-3.65 °C) and a warming trend (0.03°C/decade) described with some detail in S19 (see Fig. 15a,b and Fig. 26 of S19).

JFM time series of SST anomalies averaged over WSPNA are shown in Fig. 7b-f. The observed anomalies present distinctive decadal variations, with warm phases before 1970 and from late 1990s until about 2010, and a cold phase between 1970 and early 1990s. These decadal variations are modestly represented by ORAS5 ($r = 0.77$ and RMSE $= 0.28$°C, Fig. 7b), which has weaker and out of phase anomalies, and are better represented by the assimilation runs ($r = 0.94$ and RMSE $= 0.16$°C for the ensemble mean, Fig. 7c). Year 1 hindcasts perform modestly ($r = 0.43$ and RMSE $= 0.42$°C, Fig. 7d), and Year 2 hindcasts poorly, showing strong decadal variations that are mostly out of phase with observations ($r = -0.6$ and RMSE $= 1.04$°C, Fig. 7e). The anomalies of the simulations, which are characterized by a warming trend, are not expected to match the internally generated variability of the observed anomalies ($r = 0.42$ and RMSE $= 0.46$°C, Fig. 7f), although the latter are mostly contained within the ensemble spread. Analogous plots for the Labrador Sea (Fig. 8b-f) show disagreements also between ORAS5 and observed JFM SST anomalies, as well as for the assimilation runs, which are imprinted in the forecasts leading to unrealistic strong and out of phase variations for Year 2. The simulation ensemble present anomalies above -0.2°C during the whole time period, which are virtually unchanged in the mean at subzero temperatures prior the year 2000 as a result of excessive sea ice (S19). The poor skill in WSPNA and Labrador Sea ~~can potentially~~ are likely to impact predictions of surface climate over North America and Europe (Eade et al., 2012; Ruprich-Robert et al., 2017a), and West Africa and the Sahel (Martin and Thorncroft, 2014b; García-Serrano et al., 2015). Predictability of the tropical Atlantic (Dunstone et al., 2011), ocean heat content in the Nordic Seas, and decadal Arctic winter sea ice trends (Yeager and Robson, 2017) could also be affected.

7 Predictability and skill of surface climate on land

One of the main motivations for decadal climate prediction is the understanding that low-frequency variations in the upper-ocean heat content can influence surface climate by inducing atmospheric circulation changes both locally and remotely (Zhang and Delworth, 2006; Ruprich-Robert et al., 2017b). The expectation is that ocean model initialization will allow skilful surface climate prediction from seasons to years (Smith et al., 2007; Doblas-Reyes et al., 2011). Assessing the influence of model initialization on forecast skill can be challenging however (Smith et al., 2019), particularly in the presence of strong secular trends that can hinder the detection of internally generated predictable variations. Meehl et al. (2020) report that CanESM5 has equilibrium climate sensitivity (ECS) and transient climate response (TCR) at or near the top of the range among the Earth system models participating in CMIP6, with ECS = 5.6 K in the 1.8 K to 5.6 K range, and TCR = 2.7 K in the 1.3 K to 3.0 K range. CanESM5 also exhibits a strong historical warming trend (Figs. 25a and 26, S19), which leads to global temperatures changes exceeding those observed toward the end of the historical period, especially over the Arctic and on land regions. Therefore, improvement of forecast skill in CanESM5 SAT predictions over land can be expected not only due to the impact of initialization on internally generated variability, but also on corrections to this excessive warming trend.

Figures 9a-c show q_e of annual mean near-surface air temperature (SAT) on land for Year 2, Year 2-5 and Year 6-9 ~~forecasts~~ hindcasts. For Year 2, the *ppvf* is generally largest in the tropics ($q_e > 0.4$), where atmospheric circulation is most strongly influenced by SST (Lindzen and Nigam, 1987; Smith et al., 2012). Tropical regions are impacted by initialization ($q_{e_i} > 0.1$), most notably in the Amazon, where $q_{e_i} \approx 0.2$ to 0.4 (Fig. 9d). Extratropical regions are characterized by a relatively higher atmospheric noise, thus displaying reduced *ppvf* ($q_e < 0.4$) and little contribution from initialization. For multi-year averages, the noise component is reduced considerably leading to a relatively high *ppvf* (Fig. 9b,c), particularly in regions were the warming trend is dominant (Fig. 10). The impact of initialization is also reduced, with typically $q_{e_i} < 0.1$ (Fig. 9e,f).

SB20 show that annual and multiyear averages of CanESM5 SAT hindcasts have significant correlation skill over most land regions due to the strong temperature response to external forcing, with a modest contribution from initialization. In terms of MSSS, SAT hindcasts on land are more skilful than observed climatology (MSSS>0) in the tropics and in regions near to large water masses (Fig. 11a-c), mirroring the behaviour of q_e in Fig. 9a-c. Skill is highest for multi-year averages, where significant MSSS values are also seen inland (Fig. 11b,c). Notably, MSSS<0 in the Amazon despite of positive correlation skill (Fig. 3 of SB20), indicating an excessive variance in the hindcasts possibly due to unrealistic trends (Fig. 10). The hindcasts outperform persistence in the extratropics, but underperform in the tropics most notably in Year 2 (Fig. 11d), primarily over regions of unrealistic trends (Fig. 10). By contrast, the hindcasts outperform simulations in the tropics (Fig. 11g) where initialization contribute to *ppvf* (Fig. 9d) and correlation skill (Fig. 3e of SB20), except for central Africa and the Sahel. For multi-year averages, the hindcasts outperform simulations in most regions (Fig. 11h,i) despite little impact from initialization to correlation skill (Fig. 3f of SB20), suggesting that the improvements are likely due to reductions of the simulated trends (Fig. 10).

The upper panels of Fig. 12 show q_e of annual mean precipitation hindcasts for Year 2, Year 2-5 and Year 6-9. For Year 2, $q_e > 0.1$ is confined to tropical and subtropical regions, with slightly higher values in the Amazon basin ($0.2 < q_e < 0.3$). The precipitation signal extends to higher latitudes and is relatively stronger for multi-year averages. The largest *ppvf* values are seen in the Sahel for longer lead times ($q_e > 0.5$ for Year 6-9 in some locations) as the externally forced component becomes dominant. Generally, most of the hindcast precipitation signal with $q_e > 0.1$ is externally forced. The largest contribution of initialization to *ppvf* are seen in northeast Brazil, central Southwest Asia, and southern Australia for Year 2 hindcasts ($0.1 < q_e < 0.2$), and $q_{e_i} < 0.1$ elsewhere (Fig. 12d-f). The negative values of q_{e_i} seen in the plots are the result of a negligible impact from initialization and sampling errors. Sources of the predictable signal and prediction skill in the northeast Brazil, central Southwest Asia and the Sahel are discussed in section 8.

The correlation skill of the annual mean precipitation hindcasts (Fig. 13a-c) partly mirrors the patterns of Fig. 12a-c, but can be significant also in regions of little *ppvf*. Correlation skill tends to increase both in magnitude and geographic extent for multi-year averages (Fig. 13b,c). A large component of skill is attributed to the uninitialized forced component, as can be inferred from Figs. 13d-f. Known sources of externally forced decadal precipitation variability include drivers of climate change such as CO_2 and anthropogenic SO_2, which can alter the energy budget due to changes in the atmospheric composition leading to climate feedback processes that affect precipitation (Myhre et al., 2017). Another major source is volcanic aerosols, which are injected into the stratosphere during a volcanic eruption and can reduce global mean temperature leading to a dryer atmosphere and reduced precipitation 2 to 3 years after an event (Smith et al., 2012). This is shown in Fig. 14 by the difference of mean precipitation from hindcasts with and without volcanic forcing following three major volcanic events (Agung, 1963; El Chichón, 1982 and Pinatubo, 1991). The setup of these volcanic experiments are briefly described in section 3 and follows the recommendations of ~~by~~ Boer et al. (2016) as part of Component C of the DCPP. The difference in mean precipitation is significant over various land regions and most notably in the Maritime Continent. Precipitation ~~forecasts~~ hindcasts over this region have significant correlation skill (Fig. 13a-c) not ~~robustly~~ fully associated to initialization (Fig. 13d-f), thus volcanic forcing ~~seems to~~ could be a contributing factor. Contributions from initialization to correlation skill are significant in a few regions including northeast Brazil, western North America and central Southwest Asia for Year 2 and Year 2-5 (Figs. 13d,e), and are much reduced for Year 6-9 (Figs. 13f).

8 Skill dependence on ensemble size

Large single- and multi-model ensembles of initialized and uninitialized predictions have become essential tools in the study of decadal climate predictions due in part to the considerable noise reduction that can be achieved by ensemble averaging (Yeager et al., 2018; Deser et al., 2020; Smith et al., 2020). ~~For a single-model ensemble~~Generally, the ensemble size required to extract predictable signals varies among climatic variables, and may depend on forecast range, in the case of initialized predictions, and on geographic location. The tendency for models to underestimate predictable signals (Scaife and Smith, 2018; Smith et al., 2020) reinforces the need for large ensembles.

~~The noise-to-signal variance ratio of forecast and simulation ensembles have the form $\gamma_Y = \sigma^2_{y_e}/\sigma^2_{\Psi}$, Eq. (A12), and $\gamma_U = \sigma^2_{u_e}/\sigma^2_{\Phi}$, Eq. (A13), respectively. Figure 15 plots γ_Y and γ_U of annual mean and multi-year terrestrial precipitation for the 40-member ensembles.~~ Figure 15 shows the noise-to-predictable variance ratio γ_Y, Eq. (A12), and γ_U, Eq. (A13), of annual and multi-year mean terrestrial precipitation for the 40-member ensembles of hindcasts and uninitialized simulations. The global land averages of γ_Y and γ_U as a function of ensemble size stabilize for $\gtrsim 10$ members (not shown), suggesting that the patterns of Fig. 15 are largely robust under changes in ensemble samples and sizes. In terms of $q = \sigma^2_{\Psi}/\sigma^2_Y$, the ensemble size required for ~~$q > \alpha$~~ $q > q_0$ is, according to Eq. (5), ~~$m_Y > \gamma_Y \alpha (1-\alpha)^{-1}$~~ $m_Y > \gamma_Y q_0 \left(1 - q_0\right)^{-1}$, so ~~$q > 0.9$~~ $q > q_0 = 0.9$ requires $m_Y > 9\gamma_Y$. Therefore, all regions in Fig. 15a,b with, say, $\gamma_Y > 5$ require $m_Y > 45$ members to satisfy $q > 0.9$, i.e., over 45 members are needed for the variance of the ensemble mean ~~forecast~~ hindcast to be at least 90% predictable. Most regions in Fig. 15a,b have $\gamma_Y > 5$, suggesting a benefit of large ensembles. A few exceptions include the Amazon basin for inter-annual ~~time scales~~ variations (Fig. 15a), and the Sahel for multi-year ~~averages~~ variations (Fig. 15b), which are both characterized by relatively strong precipitation signals. A similar analysis can be made for the uninitialized simulations.

To illustrate forecast skill dependence on ensemble size, precipitation predictions over northeast Brazil (NEB; 10°S-5°N, 50°W-35°W) and central Southwest (CSW) Asia (25°N-55°N, 40°E-75°E) are considered. These two regions stand out for the potentially predictable precipitation signal (Fig. 15a,b) and associated correlation skill due to initialization (Fig. 13a,d). Precipitation variability over NEB has been linked to variations of the inter-tropical convergence zone modulated by Atlantic SST gradients and tropical Pacific SST anomalies, the latter mainly driven by El Niño Southern Oscillation on interannual time scales, which are in turn modulated by the Atlantic Multidecadal Variability and the Interdecadal Pacific Oscillation on decadal time scales (Nobre et al., 2005; Villamayor et al., 2018). Over CSW Asia, wintertime precipitation anomalies have been linked to variations of the East Asian jet stream driven partly by western Pacific convection and SST anomalies, and Maritime Continent convection (Barlow et al., 2002; Tippett et al., 2003).

Figures 16a,b show the dependence on ensemble size of the variance contributions $q_u = \sigma^2_{Y_u}/\sigma^2_Y$ and $q_i = \sigma^2_{Y_i}/\sigma^2_Y$ to correlation skill r_{XY} from the uninitialized Y_u and initialized Y_i components of Year 2 annual mean precipitation ~~forecasts~~ hindcasts averaged over NEB and CSW Asia. For both regions, $q_u \lesssim 0.2$ for all ensemble sizes indicating small variance contribution to skill from the simulated response to external forcing. By contrast, q_i increases from about 0.40 for ensemble size m_Y =10 to about 0.65 for $m_Y = 40$ over NEB, and from about 0.50 for m_Y =10 to about 0.80 for $m_Y = 40$ over CSW Asia, showing that initialization impacts the *ppvf* $q = q_i + q_u$ in Eq. (5), and that large ensembles are required to extract the initialized predictable variance from the ensemble mean ~~forecast~~ hindcast. The variance contribution to correlation skill will increase further, albeit minimally and slowly, for ensemble sizes larger than 40, so there is a limit to the cost-effective increase of ensemble size to improve skill. For Year 2-5 the behaviour is somewhat similar (Fig. 17a,b) although the variance contribution of the initialized (uninitialized) component tends to be lower (higher).

Besides their variance contribution to skill, Y_u and/or Y_i must have realistic variations in phase for a skilful ensemble mean prediction Y. The correlation skill r_{XY} for Year 2 annual mean precipitation ~~forecast~~ hindcast averaged over NEB and CSW Asia is shown in Figs. 16c,d as a function of ensemble size. Also shown are the correlation skills r_{XU} of the ~~historical~~ uninitialized simulations and r_{XX_p} of the persistence forecast. For both regions, the ~~forecast~~ hindcast correlation skill is positive at the 90% confidence level. ~~Forecasts~~ Hindcasts skill increases with ensemble size and surpasses that of uninitialized simulations for $m_Y \gtrsim 15$, indicating an added value from initialization that would have been underestimated for $m_Y < 15$ by this metric. Unlike uninitialized simulations, the ~~forecasts~~ hindcasts over NEB are more skilful than persistence for all ensemble sizes (Fig. 16c). By contrast, the median correlation skill over CSW Asia surpasses persistence for $m_Y \gtrsim 20$, but may require more than 40 members to do so confidently (Fig. 16d). It should be noted, however, that forecast correlation skill is higher when averaged over winter and spring (DJFMAM), and surpasses that of persistence and of the simulations for $m_Y > 10$ (not shown). This is consistent with the seasonal cycle of mean precipitation over CSW Asia (Tippett et al., 2003; Schiemann et al., 2008), as the precipitation signal is stronger during DJFMAM. For Year 2-5 (Figs. 17c,d), the ~~forecasts~~ hindcasts over NEB are more skilful than uninitialized simulations for $m_Y \gtrsim 20$, but require $m_Y \gtrsim 35$ to marginally outperform simulations over CSW Asia, indicating an advantage of large ensembles.

The results over the Sahel are somewhat different. The Sahel is an important benchmark for the assessment of decadal predictions due to its strong summer rainy season, the variation of which is considered one of the largest signals of global climatic variability on annual to multi-year time scales (Martin and Thorncroft, 2014a; Sheen et al., 2017). Previous studies indicate that initialization enhances the skill of Sahelian rainfall predictions compared to simulations, although results vary among models (Garcia-Serrano et al., 2013; Gaetani and Mohino, 2013; Martin and Thorncroft, 2014a; Sheen et al., 2017; Yeager et al., 2018). Figures 18a,b show the dependence on ensemble size of Year 2 and Year 2-5 forecast correlation skill r_{XY} for July-August-September (JAS) mean precipitation averaged over the Sahelian sector (10°N-20°N, 20°W-10°E), as well as r_{XU} and r_{XX_p} for the simulations and persistence, respectively. Generally, ~~forecasts~~ hindcasts and uninitialized simulations outperform persistence by a large margin, but both exhibit about the same level of skill suggesting virtually no impact from initialization. The increase in skill is accompanied by a reduction in skill uncertainty, illustrating a benefit of large ensembles. The correlation skill decomposition indicates that the externally forced component is the main contributor to forecast skill with a negligible impact from initialization (Fig. 18c,d).

The small impact of initialization on Sahelian rainfall ~~forecasts~~ hindcasts is at odds with previous findings (Gaetani and Mohino, 2013; Yeager et al., 2018). Inter-annual and multidecadal variability of Sahelian rainfall has been linked to SST variability in the global ocean (Rowell et al., 1995), the Atlantic (Ward, 1998; Knight et al., 2006; Zhang and Delworth, 2006; Ting et al., 2009; Martin and Thorncroft, 2014a, b; Yeager et al., 2018), the Pacific and Indian oceans (Mohino et al., 2011b; Sheen et al., 2017), and the Mediterranean sea (Rowell, 2003; Mohino et al., 2011a; Sheen et al., 2017). Greenhouse gases and aerosols have also been linked to decadal variability and trends of Sahelian rainfall by their impact on Atlantic inter-hemispheric SST gradients and resulting effect on the intertropical convergence zone (Biasutti and Giannini, 2006; Haywood et al., 2013; Hua et al., 2019; Bonfils et al., 2020), and by a direct effect of changes in radiative forcing (Haarsma et al., 2005; Biasutti, 2013; Dong and Sutton, 2015). Despite the negligible impact from initialization, CanESM5 precipitation skill over the Sahel is relatively high, particularly for Year 2-5 ($r_{XY} \approx 0.7$, Fig. 18b) and other multi-year averages (not shown), and comparable to the skill of CMIP5/6 decadal predictions from other models (Gaetani and Mohino, 2013; Yeager et al., 2018). This may be an indication that at least part of the enhanced forecast prediction skill of Sahelian rainfall in some CMIP5/6 models relative to that of uninitialized simulations might be a consequence of the impact of initialization on the forced component rather than a skilful prediction of the internally generated variability itself (i.e., due to the term in parenthesis in the definition of Y_i in Eq. (3)).

The ratio Π_r, Eq. (8), for the Sahelian JAS mean precipitation ~~forecasts~~ hindcasts increases with ensemble size and confidently surpasses 1 with 40 members for Year 2 (Fig. 18e) and approximately 15 members for Year 2-5 (Fig. 18f), indicating that for larger ensembles $r_{XY} > \rho$, i.e., the ensemble mean hindcast~~s~~ is more skilful at predicting the observed climate system than the hindcasts themselves. This is a consequence of the noise-to-predictable variance fraction γ_Y being too high, suggesting that the hindcasts are either too noisy or its predictable component is too weak relative to the observed precipitation signal. Because the hindcasts and observed total JAS precipitation variances are about the same (not shown), we conclude that the ensemble mean hindcast underestimates the amplitude of the predictable precipitation signal in the Sahel. A similar behaviour is seen for the simulations with a somewhat reduced predictable variance fraction for large ensembles, particularly for Year 2-5 (Fig. 18f), which is primarily a consequence of the stronger potentially predictable variance of the simulations (not shown). Such a behaviour is not specific to CanESM5, nor to the region, climate variable and time scales involved. It is a feature across model simulations of various climate phenomena (Scaife and Smith, 2018; Yeager et al., 2018; Smith et al., 2020), pointing to model deficiencies at representing the strength of predictable signals of the climate system.

9 Aspects of the skill of land and ocean biogeochemistry

CanESM5 models the ~~interaction between~~ effects of the physical climate~~,~~ on the biosphere~~,~~ and the chemical constituents of the atmosphere and ocean. This enables the assessment of some aspects of the predictability of ocean and land biogeochemistry, and the carbon cycle. Gross primary productivity (GPP) is the rate of photosynthetic carbon fixation by primary producers, such as phytoplankton in the ocean and plants on land. GPP of terrestrial vegetation is a key variable of the global carbon cycle and is an important component of climate change (Zhang et al., 2017). Net primary productivity (NPP) is the difference between GPP and the fraction of fixed carbon that primary producers use for respiration (Gough, 2011; Sigman and Hain, 2012), and is thus a major determinant of carbon sinks and a key regulator of ecological processes (Field et al., 1998). The potential for predictive skill of NPP hindcasts in the ocean and GPP hindcasts on land is assessed here by correlation with the assimilation runs used for initialization. We also show preliminary comparisons with observation-based estimates, but do not provide a full assessment of actual skill due to the relatively short time span of the observations. As in previous sections, uninitialized simulations are used here as a reference to quantify the impact of initialization on hindcast correlation skill. We emphasize that there is no assimilation of observed carbon cycle variables to initialize the hindcasts (section 3), therefore initial variations of GPP and NPP are the result of ensemble spread of oceanic and atmospheric states in the assimilation runs.

Figure 19 shows the correlation skill r_{XY} and the contribution from initialization r_i of ocean NPP for Year 2, Year 2-5 and Year 6-9 ~~forecasts~~ hindcasts relative to the assimilation runs. For Year 2, there is significant correlation skill in most of the global ocean north of the Antartic Circumpolar Current, except for scattered regions including WSPNA, the western North Pacific and, to some degree, in the eastern equatorial regions of the Pacific and Atlantic oceans. These regions are characterized by relatively low prediction skill of upper-ocean heat content (Fig. 2a). High NPP correlation skill is found in most ocean eastern boundaries and coastal upwelling regions, and in broader sectors associated with boundary currents including the North Pacific and California Currents, the Gulf Stream, North Atlantic and Canary Currents, the Brazil and Benguela Currents, the Agulhas Current, the East Australian Current, and in areas of the Arabian Sea and Bay of Bengal north of the Monsoon Drift. Part of this skill is attributed to initialization (Fig. 19d) with little or no impact from the simulated external forcing. Predictive skill tends to be larger in both magnitude and extent for Year 2-5 (Fig. 19b,e), and is much reduced for Year 6-9 (Fig. 19c,f) except for a few regions of relatively high skill including major eastern boundary upwelling systems (EBUS; Chan, 2019). EBUS comprise some of the ocean's most productive regions supporting approximately one-fifth of the world's ocean wild fish harvests (Pauly and Christensen, 1995) and the habitats for multiple species of pelagic fish, migratory seabirds, and marine mammals (Block et al., 2011), thus the potential for NPP skilful forecasts in these regions at relatively long lead times may have useful implications for fisheries and environmental managers. Preliminary comparisons against observation-based estimates over the Canary Current region (Fig. 21a), which along with the California, Humboldt and Benguela current systems is one of the four major EBUS (Gómez-Letona et al., 2017), show realistic inter-annual variations in the assimilation runs and Year 1 hindcasts. Ilyina et al. (2020) point out difficulties however in CanESM5 predictions of ocean CO_2 uptake in an inter-comparison of ~~e~~Earth system model results.

On land, significant GPP correlation skill of annual and multi-year hindcasts relative to the assimilation runs is found on all continents (Fig. 20a-c), although negative skill can be seen mainly in grassland and savanna regions of South America, Africa and east Asia. Correlation skill is highest in the temperate zone of eastern North America, in South East Asia and the Maritime Continent, in sectors of tropical South America and Africa, in Southern Australia, and in North Africa, the Nile basin and Arabian Peninsula. Except for the latter, these regions are characterized by moderate to high annual mean primary productivity (Fig. 1 of Field et al., 1998). Unlike ocean NPP, a large portion of GPP skill on land derives from the simulated externally forced component, particularly from CO_2 fertilization, with a moderate but significant contribution from initialization (Fig. 20d,e). The effects of initialization become negligible for longer forecast ranges, except for a small sector of the Amazon rainforest (Fig. 20f). Preliminary comparisons against observation-based products show realistic global mean GPP anomaly trends (not shown) and interannual variations for the assimilation runs and Year 1 hindcasts (Fig. 21b). This is consistent with multi-model comparisons (Ilyina et al., 2020) showing significant actual correlation skill of CO_2 land uptake in linearly detrended CanESM5 assimilation runs and hindcasts for up to 2 years. Comparisons against observation-based data however are limited by the relatively short time span and uncertainty of the observations.

10 Summary and conclusions

CanESM5 decadal hindcasts, which are CCCma's contribution to Component A of the DCPP component of CMIP6, have the ability to represent realistic inter-annual and multi-year variations of key physical climate fields and carbon cycle variables on decadal time scales. The hindcasts are 40 ensemble members retrospective forecasts that are full-field initialized from realistic climate states at the end of every year during 1960~~–2016~~ to present and run for 10 years. Natural and anthropogenic external forcing associated with greenhouse gases and aerosols are specified, and a 40-member ensemble of historical uninitialized climate simulations with the same external forcing is also produced. The predictable component of the simulations is determined by the model's response to external forcing, whereas the forecasts have predictable components due to both the initialization of internal climate states and to the model's response to external forcing, which is generally different from that of simulations. The decomposition of the predictable component of the forecasts into initialized and uninitialized constituents, the latter derived from the projection of the forecasts responses to external forcing onto that of simulations, allowed the quantification of the impact of initialization on skill, and sheds new light on the value added by a forecasting system over that of climate simulations.

The upper-ocean heat content of CanESM5 is shown to be potentially predictable during the 10-year forecast range most notably in the extratropics, with potentially predictable variance in the eastern ocean boundaries for up to the 2- to 4-year range as a result of initialization. The hindcasts realize some of this potential predictability and have actual skill largely driven by external forcing, with significant contributions from initialization in the Pacific and Indian ocean basins. Sea surface temperature (SST) ~~forecasts~~ hindcasts are skilful for most of the global ocean mainly due to the strong warming response in the model, with a moderate impact from initialization to correlation skill beyond the first year of the ~~forecasts~~ hindcasts. Compared to heat content, SST is more directly affected by atmospheric conditions reducing the contribution of initialization to skill. Initialization also decreases MSE significantly relative to that of simulations in the northern subtropics and in the Indian Ocean due to a reduction of the simulated warming trend, which highlights the impact of initialization not only on the predictability of internal climate variations, but also on corrections of the simulated response to external forcing.

The western subpolar North Atlantic (WSPNA) and the Labrador Sea regions stand out for the negative skill of the upper-ocean heat content and the surface temperature, resulting in part from erroneous temperature and salinity trends in the reanalysis data used to initialize the forecasts. Winter SST variations of CanESM5 hindcasts in these regions have strong decadal variations that are out of phase with observations beyond the 1-year range. Also, strong cold biases and warming trends in the simulations contribute to the poor performance in these regions. The lack of skill in the WSPNA and the Labrador Sea merit further analysis as it may impact climate predictability elsewhere.

The strong warming response of CanESM5 drives the potential predictability of near-surface air temperature over land, and is largely responsible for the ~~forecast~~ hindcast correlation skill as examined in SB20. Initialization, however, reduces the strength of the model response to external forcing leading to a lower ~~forecast~~ hindcast MSE than that of the simulations and persistence at all forecast ranges considered, except for some tropical regions. The correlation skill of annual and multi-year mean precipitation is, perhaps surprisingly, very high in vast continental regions including Siberia, central Southwest Asia, Northeast Europe, the Americas, and the Sahel. The precipitation skill is mainly driven by external forcing, with a non-negligible impact from volcanic aerosols, although long-lived effects from initialization can be seen in regions such as Northeastern Brazil and central Southwest Asia which can be influenced by remote SST anomalies. Skill tends to be highest for multi-year averages, as potentially erroneous inter-annual variability is averaged out and the forced component becomes dominant.

Two additions to CCCma's contribution to the decadal prediction component of CMIP6 compared to CMIP5 are the increased ensemble size to 40 members from 10 members, and the inclusion of ~~the~~ carbon cycle variables in these experiments.

There is a growing evidence that large ensemble sizes are advantageous for decadal predictions, and this work is consistent with that view. Skilful CanESM5 precipitation ~~forecasts~~ hindcasts with a significant impact from initialization require large ensembles to confidently surpass the skill of uninitialized simulations, compared to 10 or fewer members. There is however a limit to the cost-effective increase of ensemble size needed to improve skill, which is determined by the ensemble forecast noise-to-predictable variance ratio. Large ensembles are also used to show that CanESM5 decadal hindcasts underestimate the inter-annual and multi-year Sahelian summer rainfall signal, an important benchmark for the assessment of decadal predictions, as correlation skill is larger than potential correlation skill for sufficiently large ensembles despite the hindcasts having realistic total precipitation variance in this region. ~~Initialized~~ CanESM5 decadal hindcasts are skilful compared to assimilated values for predictions of net primary productivity in the ocean northward of the Antarctic Circumpolar Current for the 2- to 4-year range, with regions of long-lived skill encompassing the 10-year forecast range. A significant portion of this skill is attributed to initialization, particularly in major eastern boundary upwelling systems where there is indication of actual skill as well, and in the Bay of Bengal. On land, gross primary productivity hindcasts ~~are potentially skilful~~ have potential for skill at all ranges examined, mostly because of the CanESM5 response to the externally forced CO_2 increase, with a moderate but significant short-lived impact from initialization. Preliminary comparisons of CanESM5 assimilation runs and Year 1 ~~forecasts~~ hindcasts with observation-based products have shown agreement in the global mean anomaly trend and interannual variations for the years of available data. A comprehensive assessment of actual skill remains however a challenge due the relatively short time span and uncertainty of the verifying observations.

Code and data availability. The CanESM5 source code is publicly available at https://gitlab.com/cccma/canesm. The version of the code used to produce all the simulations described in this paper, which are submitted to CMIP6, is tagged as v5.0.3 with associated DOI: https://doi.org/10.5281/zenodo.3251113 (Swart et al., 2019a, b). The CanESM5 data for the decadal experiments (Sospedra-Alfonso et al., 2019a, b, c, d, e, f, g, h) and historical simulations (Swart et al., 2019c) are publicly available from the Earth System Grid Federation (https://esgf-node.llnl.gov/search/cmip6/). The observation-based products used here are freely available. Further details are given in appendix B.

Author contributions. RSA initiated the study, helped to coordinate and performed model simulations, produced the analysis and figures, and wrote the paper. WJM led the experimental design and development of the CCCma contribution to DCPP, and contributed with the analysis and interpretation of results. GJB was instrumental to the CCCma contribution to DCPP, and contributed with the methodology, analysis and interpretation of results. VVK and WL setup and produced the bulk of model simulations. CS contributed with results, plotting, and assessment of GPP on land. JRC contributed with results and assessment of NPP in the ocean. All authors contributed with the manuscript and the final version of the paper.

Competing interests. The authors declare that they have no conflict of interests.

Acknowledgements. We are thankful to the CCCma team at large for the development and production of the CanESM5 ~~e~~Earth system model. We are thankful also to Neil Swart and Michael Sigmund for their helpful comments on an earlier version of this paper. We acknowledge the various organizations listed in table B2 for making the verification data used here freely available. This work is dedicated to the memory of our colleague and friend Fouad Majaess.

Appendix A: Details of hindcast evaluation methods

A1 Associated variances

The variances associated to the ensembles of forecasts and simulations in Eq. (1), together with those of the ensemble mean of forecasts and simulations in Eq. (2), are

$$
\begin{aligned}
\sigma^2_{Y_e} &= \sigma^2_\Psi + \sigma^2_{y_e} = \sigma^2_{\psi_f} + \sigma^2_\psi + \sigma^2_{y_e} && \text{(A1)}\\
\sigma^2_{U_e} &= \sigma^2_\Phi + \sigma^2_{u_e} = \sigma^2_{\phi_f} + \sigma^2_{u_e} && \text{(A2)}\\
\sigma^2_Y &= \sigma^2_\Psi + \sigma^2_y = \sigma^2_\Psi + \frac{\sigma^2_{y_e}}{m_Y} = \sigma^2_{Y_u} + \sigma^2_{Y_i} + \frac{\sigma^2_{y_e}}{m_Y} \longrightarrow \sigma^2_{Y_u} + \sigma^2_{Y_i} && \text{(A3)}\\
\sigma^2_U &= \sigma^2_\Phi + \sigma^2_u = \sigma^2_{\phi_f} + \frac{\sigma^2_{u_e}}{m_U} \longrightarrow \sigma^2_{\phi_f} && \text{(A4)}
\end{aligned}
$$

where we have used $\sigma^2_y = \sigma^2_{y_e}/m_Y$ under the assumption of the y_k's independence, and similarly for u_k. The noise variances are estimated from Eqs. (A1)-(A4) as

$$
\begin{aligned}
\sigma^2_{y_e} &= \frac{m_Y}{m_Y - 1}\left(\sigma^2_{Y_e} - \sigma^2_Y\right) && \text{(A5)}\\
\sigma^2_{u_e} &= \frac{m_U}{m_U - 1}\left(\sigma^2_{U_e} - \sigma^2_U\right) && \text{(A6)}
\end{aligned}
$$

whereas the predictable variances are estimated from Eqs. (A3)-(A6) as

$$
\begin{aligned}
\sigma^2_\Psi &= \sigma^2_Y - \frac{\sigma^2_y}{m_Y} = \frac{m_Y \sigma^2_Y - \sigma^2_{Y_e}}{m_Y - 1} && \text{(A7)}\\
\sigma^2_\Phi &= \sigma^2_U - \frac{\sigma^2_u}{m_u} = \frac{m_U \sigma^2_U - \sigma^2_{U_e}}{m_U - 1} && \text{(A8)}
\end{aligned}
$$

The predictable and noise variances can be readily computed from the data by means of the total variance $\sigma^2_{Y_e}$ or $\sigma^2_{U_e}$, and the variance of the ensemble mean σ^2_Y or σ^2_U. If we write explicitly the dependence of the anomaly forecast $Y_{jk}(\tau)$ and ensemble mean $Y_j(\tau)$ on the forecast range τ, ensemble member $k = 1 \ldots m_Y$, and initial year $j = 1 \ldots n$, then

$$
\begin{aligned}
\sigma^2_{Y_e}(\tau) &= \frac{1}{m_Y\,(n-1)} \sum_{j=1}^{n} \sum_{k=1}^{m_Y} \left[Y_{jk}(\tau) - \overline{Y}_k(\tau)\right]^2 && \text{(A9)}\\
\sigma^2_Y(\tau) &= \frac{1}{(n-1)} \sum_{j=1}^{n} \left[Y_j(\tau) - \overline{Y}(\tau)\right]^2 && \text{(A10)}
\end{aligned}
$$

and similarly for the simulations, where the over line indicates the average over the initial years.

A2 Correlation skill decomposition

Following SB20, the correlation skill of the ensemble mean forecast can be decomposed as

$$r_{XY} = r_{XY_u}\frac{\sigma_{Y_u}}{\sigma_Y} + r_{XY_i}\frac{\sigma_{Y_i}}{\sigma_Y} = r_u + r_i \tag{A11}$$

where r_{XY_u} and r_{XY_i} are the correlation skills of the uninitialized and initialized components Y_u and Y_i themselves, while r_u and r_i are the components contribution when scaled by the fractions of the variances involved. In terms of the noise-to-predictable variance ratios of forecasts and simulation,

$$\gamma_Y = \frac{\sigma^2_{y_e}}{\sigma^2_\Psi} = \frac{m_Y\left(\sigma^2_{Y_e} - \sigma^2_Y\right)}{m_Y\sigma^2_Y - \sigma^2_{Y_e}} \tag{A12}$$

$$\gamma_U = \frac{\sigma^2_{u_e}}{\sigma^2_\Phi} = \frac{m_U\left(\sigma^2_{U_e} - \sigma^2_U\right)}{m_U\sigma^2_U - \sigma^2_{U_e}} \tag{A13}$$

and available correlations and variances, these quantities can be computed explicitly as

$$\frac{\sigma^2_{Y_u}}{\sigma^2_Y} = \theta r^2_{YU}\left(1+\frac{\gamma_U}{m_U}\right) \longrightarrow \theta r^2_{YU} \tag{A14}$$

$$\frac{\sigma^2_{Y_i}}{\sigma^2_Y} = -\theta r^2_{YU}\left(1+\frac{\gamma_U}{m_U}\right) + \left(1+\frac{\gamma_Y}{m_Y}\right)^{-1} \longrightarrow 1-\theta r^2_{YU} \tag{A15}$$

$$r_u = \theta r_{XU} r_{YU}\left(1+\frac{\gamma_U}{m_U}\right) \longrightarrow \theta r_{XU} r_{YU} \tag{A16}$$

$$r_i = r_{XY} - \theta r_{XU} r_{YU}\left(1+\frac{\gamma_U}{m_U}\right) \longrightarrow r_{XY} - \theta r_{XU} r_{YU} \tag{A17}$$

$$r_{XY_u} = \theta r_{XU}\left(1+\frac{\gamma_U}{m_U}\right)^{1/2} \longrightarrow \theta r_{XU} \tag{A18}$$

$$r_{XY_i} = \frac{\sigma_Y}{\sigma_{Y_i}}\left[r_{XY} - \theta r_{XU} r_{YU}\left(1+\frac{\gamma_U}{m_U}\right)\right] \longrightarrow \left(1-\theta r^2_{YU}\right)^{-1/2}\left[r_{XY} - \theta r_{XU} r_{YU}\right] \tag{A19}$$

where r_{YU} denotes the correlation between the ensemble means of forecasts and simulations, and the step function $\theta = 0$ if $r_{YU} < 0$, else $\theta = 1$. The ratios γ_Y and γ_U are estimated according to Eq. (A12) and Eq. (A13) with the total variances $\sigma^2_{Y_e}$ and $\sigma^2_{U_e}$, Eq. (A9), and the ensemble mean variances σ^2_Y and σ^2_U, Eq. (A10), for simulations and forecasts. For finite ensembles, σ^2_U and σ^2_Y and thus γ_U and γ_Y can be negative due to sampling errors. With Eqs. (A12)-(A13), the quantities in Eqs. (A16)-(A19) are readily computed from the data.

Appendix B: Data sources, variables, and derived quantities

Table B1. List of figures, CMIP6 variables, experiments, and verifying observation-based products employed in this paper. See table B2 for the sources of the observation-based products. The entry "n/a" indicates "not applicable".

Figure	CMIP6 variable	CMIP6 experiment and variant label	Observation-based product
1	thetao (vertically integrated in the upper 300m)	dcppA-hindcast, r[1-40]i1p2f1	n/a
2	thetao (vertically integrated in the upper 300m)	dcppA-hindcast, r[1-40]i1p2f1	EN4.2.1
3	tos	dcppA-hindcast, r[1-40]i1p2f1	n/a
4	tos	dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	ORAS5
5, 6	tos	dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	ERSSTv5
7, 8	tos	dcppA-assim, r[1-10]i1p2f1 dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	ERSSTv5, ORAS5
9	tas	dcppA-hindcast, r[1-40]i1p2f1	n/a
10, 11	tas	dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	ERA-40, ERA-Interim
12	pr	dcppA-hindcast, r[1-40]i1p2f1	n/a
13	pr	dcppA-hindcast, r[1-40]i1p2f1	GPCP2.3
14	pr	dcppC-hindcast-noAgung, r[1-10]i1p2f1 dcppC-hindcast-noElChichon, r[1-10]i1p2f1 dcppC-hindcast-noPinatubo, r[1-10]i1p2f1 dcppC-forecast-addAgung, r[1-10]i1p2f1 dcppC-forecast-addElChichon,r[1-10]i1p2f1 dcppC-forecast-addPinatubo, r[1-10]i1p2f1	n/a
15	pr	dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	n/a
16, 17, 18	pr	dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	GPCP2.3
19, 20	intpp, gpp	dcppA-assim, r[1-10]i1p2f1 dcppA-hindcast, r[1-40]i1p2f1 historical, r[1-40]i1p2f1	n/a
21	intpp, gpp	dcppA-assim, r[1-10]i1p2f1 dcppA-hindcast, r[1-40]i1p2f1	VGPM, MODIS, GOSIF

Table B2. List of observation-based products.

Observation-based product	Citation
EN4.2.1	Met Office Hadley Centre (Good et al., 2013)
	https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html
ERSSTv5	US National Oceanic and Atmospheric Administration (NOAA)
	Extended Reconstructed Sea Surface Temperature (Huang et al., 2017)
	https://www1.ncdc.noaa.gov/pub/data/cmb/ersst/v5/netcdf
ERA40, ERA-Interim	European Centre for Medium-Range Forecasts (ERA40; Uppala et al., 2005)
	and (ERA-Interim; Dee et al., 2011)
	https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets
GPCP2.3	Global Precipitation Climatology Project (Adler et al., 2003)
	https://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html
MODIS	NASA's MODIS-based gross primary productivity product (Zhang et al., 2017)
	https://figshare.com/articles/dataset/Monthly_GPP_at_0_5_degree/5048011
GOSIF	Orbiting Carbon Observatory-2 (OCO-2)-based Solar-induced chlorophyll
	fluorescence (SIF) product (Li and Xiao, 2019)
	http://data.globalecology.unh.edu/data/GOSIF-GPP_v2
ERA5	Hersbach and coauthors (2020). Copernicus Climate Change Service (C3S)
	https://cds.climate.copernicus.eu/cdsapp#!/home
SSI	Surface solar irradiance (Bishop et al., 1997)
SeaWiFS	NASA Goddard Space Flight Center
	Sea-viewing Wide Field-of-view Sensor Chlorophyll Data; reprocessing version 2010.0
	https://oceancolor.gsfc.nasa.gov/data/10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018/
MODIS-Terra	NASA Goddard Space Flight Center
	Moderate-resolution Imaging Spectroradiometer Terra Chlorophyll Data; reprocessing version 2010.0
	https://oceancolor.gsfc.nasa.gov/data/10.5067/TERRA/MODIS/L3M/CHL/2018/
VGPM	Vertically generalized production model (Behrenfeld and Falkowski, 1997)
	Uses ocean chlorophyll from SeaWiFS (1998-2004) and
	MODIS-Terra (2005-2012), SST from ERA5, and SSI monthly climatology (1983-1991)

References

Adler, R. F., Huffman, G., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., and Arkin, P.: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present), Journal of Hydrometeorology, 4, 1147–1167, 2003.

Arora, V. K.: Simulating energy and carbon fluxes over winter wheat using coupled land surface and terrestrial ecosystem models, Agricultural and Forest Meteorology, 118, 21–47, 2003.

Arora, V. K. and Boer, G. J.: A representation of variable root distribution in dynamic vegetation models, Earth Interaction, 7, 1–19, 2003.

Arora, V. K. and Boer, G. J.: A parameterization of leaf phenology for the terrestrial ecosystem component of climate models, Global Change Biology, 11(1), 39–59, 2005.

Banzon, V., Smith, T. M., Chin, T. M., Liu, C., and Hankins, W.: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, Earth Syst. Sci. Data, 8, 165–176, 2016.

Barlow, M., Cullen, H., and Lyon, B. R.: Drought in Central and Southwest Asia: La Niña, the Warm Pool, and Indian Ocean Precipitation, Journal of Climate, 15, 697–700, 2002.

Behrenfeld, M. J. and Falkowski, P. G.: Photosynthetic rates derived from satellite-based chlorophyll concentration, Limnology and Oceanography, 42, 1–20, 1997.

Biasutti, M.: Forced Sahel rainfall trends in the CMIP5 archive, Journal of Geophysical Research: Atmospheres, 118, 1613–1623, 2013.

Biasutti, M. and Giannini, A.: Robust Sahel drying in response to late 20th century forcings, Geophysical Research Letters, 33, 2006.

Bishop, J. K. B., Rossow, W. B., and Dutton, E. G.: Surface solar irradiance from the International Satellite Cloud Climatology Project 1983–1991, Journal of Geophysical Research: Atmospheres, 102, 6883–6910, 1997.

Block, B. A., Jonsen, I. D., Jorgensen, S. J., Winship, A. J., Shaffer, S. A., Bograd, S. J., Hazen, E. L., Foley, D. G., Breed, G. A., Harrison, A.-L., Ganong, J. E., Swithenbank, A., Castleton, M., Dewar, H., Mate, B. R., Shillinger, G. L., Schaefer, K. M., Benson, S. R., Weise, M. J., Henry, R. W., and Costa, D. P.: Tracking apex marine predator movements in a dynamic ocean, Nature, 475, 86–90, 2011.

Boer, G. J., Kharin, V. V., and Merryfield, W. J.: Decadal predictability and forecast skill, Climate Dynamics, 41, 1817–1833, 2013.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Mod. Dev., 9, 3751–3777, 2016.

Boer, G. J., Kharin, V. V., and Merryfield, W. J.: Differences in potential and actual skill in a decadal prediction experiment, Climate Dynamics, 2019a.

Boer, G. J., Merryfield, W. J., , and Kharin, V. V.: Relationships between potential, attainable, and actual skill in a decadal prediction experiment, Climate Dynamics, 52, 4813–4831, 2019b.

Bonfils, C. J. W., Santer, B. D., Fyfe, J. C., Marvel, K., Phillips, T. J., and Zimmerman, S. R. H.: Human influence on joint changes in temperature, rainfall and continental aridity, Nature Climate Change, https://doi.org/doi.org/10.1038/s41558-020-0821-1, 2020.

Bouillon, S., Maqueda, M. M., Legat, V., and Fichefet, T.: An elastic-viscous-plastic sea ice model formulated on Arakawa B and C grids, Ocean Modelling, 27, 174–184, 2009.

Carrassi, A., Guemas, V., Doblas-Reyes, F. J., Volpi, D., and Asif, M.: Sources of skill in near-term climate prediction: generating initial conditions, Climate Dynamics, 47, 3693–3712, 2016.

Chan, F.: Ocean deoxygenation: Everyone's problem - Causes, impacts, consequences and solutions, chap. Global and regional case studies of ocean deoxygenation / Evidence for ocean deoxygenation and its patterns: Eastern Boundary Upwelling Systems, Gland, Switzerland: IUCN. xxii+562pp, 2019.

Christian, J. R., Arora, V. K., Boer, G. J., Curry, C. L., Zahariev, K., Denman, K. L., Flato, G. M., Lee, W. G., Merryfield, W. J., Roulet, N. T., and Scinocca, J. F.: The global carbon cycle in the Canadian Earth system model (CanESM1): Preindustrial control simulation, Journal of Geophysical Research, 115, G03 014, 2010.

Cole, J. and Coauthors: The Canadian Atmospheric Model version 5, Geoscientific Model Development, 2019.

Dee, P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, S., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Monge-Sanz, B., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M.: Insights from Earth system model initial-condition large ensembles and future prospects, Nature Climate Change, https://doi.org/10.1038/s41558-020-0731-2, 2020.

Dirkson, A., Merryfield, W. J., and Monahan, A. H.: Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions, Journal of Climate, 30, 1001–1017, 2017.

Doblas-Reyes, F. J., Balmaseda, M. A., Wisheimer, A., and Palmer, T. N.: Decadal climate prediction with the European Centre for Medium-Range Weather Forecasts coupled forecast system: Impact of ocean observations, Journal of Geophysical Research, 116, 1–13, 2011.

Dong, B. and Sutton, R.: Dominant role of greenhouse-gas forcing in the recovery of Sahel rainfall, Nature Climate Change, 2015.

Dunstone, N. J., Smith, D. M., and Eade, R.: Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean, Geophysical Research Letters, 38, 2011.

Eade, R., Hamilton, E., Smith, D. M., Graham, R. J., and Scaife, A. A.: Forecasting the number of extreme daily events out to a decade ahead, Jornal of Geophysical Research, 117, 1–12, 2012.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal to decadal climate predictions underestimate the predictability of the real world?, Geophysical Research Letters, 41, 5620–5628, 2014.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Fichefet, T. and Maqueda, M. M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, Journal of Geophysical Research, 102, 12 609–12 646, 1997.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P.: Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components, Science, 281, 237–240, 1998.

Gaetani, M. and Mohino, E.: Decadal Prediction of the Sahelian Precipitation in CMIP5 Simulations, Journal of Climate, 26, 7708–7719, 2013.

Garcia-Serrano, J., Doblas-Reyes, F. J., Haarsma, R. J., and Polo, I.: Decadal prediction of the dominant West African monsoon rainfall modes, Journal of Geophysical Research, 118, 5260–5279, 2013.

García-Serrano, J., Guemas, V., and Doblas-Reyes, F. J.: Added-value from initialization in predictions of Atlantic multi-decadal variability, Climate Dynamics, 44, 2539–2555, 2015.

Goddard, G. J., Kumar, A., Solomon, A., Smith, D., Boer, G., González, P., Kharin, V., Merryfield, W., Deser, C., Mason, S. J., Kirtman, B. P., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C. A. T., Stephenson, D. B., Meel, G. A., Stockdale, T., Burgman, R., Greene, A. M., Kushnir, Y., Newman, M., Carton, J., Fukumori, I., and Delworth, T.: A verification framework for interannual-to-decadal predictions experiments, Climate Dynamics, 40, 245–272, 2013.

Gómez-Letona, M., Ramos, A. G., Coca, J., and Arístegui, J.: Trends in Primary Production in the Canary Current Upwelling System—A Regional Perspective Comparing Remote Sensing Models, Frontiers Marine Science, 4, 2017.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, Journal of Geophysical Research, 118, 6704–6716, 2013.

Gough, C. M.: Terrestrial Primary Production: Fuel for Life, Nature Education Knowledge, 3, 2011.

Haarsma, R. J., Selten, F. M., Weber, S. L., and Kliphuis, M.: Sahel rainfall variability and response to greenhouse warming, Geophysical Research Letters, 32, 2005.

Haywood, J. M., Jones, A., Bellouin, N., and Stephenson, D.: Asymmetric forcing from stratospheric aerosols impacts Sahelian rainfall, Nature Climate Change, https://doi.org/DOI: 10.1038/NCLIMATE1857, 2013.

Hersbach, H. and coauthors: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hua, W., Dai, A., Zhou, L., Qin, M., and Chen, H.: An externally forced decadal rainfall seesaw pattern over the Sahel and southeast Amazon, Geophysical Research Letters, 46, 2019.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5), Upgrades, validations, and intercomparisons, Journal of Climate, 30, 8179–8205, 2017.

Ilyina, T., Li, H., , Spring, A., Müller, W. A., Bopp, L., Chikamoto, M. O., Danabasoglu, G., Dobrynin, M., Dunne, J., Fransner, F., Friedlingstein, P., Lee, W., Lovenduski, N. S., Merryfield, W., Mignot, J., Park, J., Séférian, R., Sospedra-Alfonso, R., Watanabe, M., and Yeager, S.: Predictable variations of the carbon sinks and atmospheric CO_2 growth in a multi-model framework, Geophysical Research Letters, 2020.

Jackson, L. C., Dubois, C., Forget, G., Haines, K., Harrison, M., Iovino, D., Kohl, A., Mignac, D., Masina, S., Peterson, K. A., Piecuch, C. G., Roberts, C. D., Robson, J., Storto, A., Toyoda, T., Valdivieso, M., Wilson, C., Wang, Y., , and Zuo, H.: The mean state and variability of the North Atlantic circulation: A perspective from Ocean Reanalyses, Journal of Geophysical Research: Oceans, 124, 9141–9170, 2019.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geoscientific Model Developement, 12, 1087–1117, 2019.

Knight, J. R., Folland, C. K., and Scaife, A. A.: Climate impacts of the Atlantic Multidecadal Oscillation, Geophysical Research Letters, 33, 2006.

Levitus, S., Antonov, J. I., Boyer, T. P., Garcia, H. E., and Locarnini, R. A.: EOF analysis of upper ocean heat content, 1956–2003, Geophysical Research Letters, 32, 1–4, 2005.

Li, X. and Xiao, J.: Mapping Photosynthesis Solely from Solar-Induced Chlorophyll Fluorescence: A Global, Fine-Resolution Dataset of Gross Primary Production Derived from OCO-2, Remote Sensing, 11, 2019.

Lindzen, R. S. and Nigam, S.: On the role of sea surface temperature gradients in forcing low level winds and convergence in the tropics, Journal of Atmospheric Sciences, 44, 2418–2436, 1987.

Madec, G. and Coauthors: NEMO ocean engine, version 3.4, Note du Pole de Modélisation 27, Institut Pierre-Simon Laplace, 2012.

Martin, E. R. and Thorncroft, C. D.: Sahel rainfall in multimodel CMIP5 decadal hindcasts, Geophysical Research Letters, 41, 2169–2175, 2014a.

Martin, E. R. and Thorncroft, C. D.: The impact of the AMO on the West African monsoon annual cycle, Quarterly Journal of the Royal Meteorological Society, 140, 31–46, 2014b.

McKinnon, K. A., Rhines, A., Tingley, M. P., and Huybers, P.: Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures, Nature Geoscience, 9, 389–394, 2016.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., and Stockdale, T.: Decadal prediction: Can it be skillful, Bulletin of the American Metheorological Society, 90, 1467–1485, 2009.

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., G. J. van Oldenborgh, G. V., and Yeager, S.: Decadal climate prediction: An update from the trenches, Bulletin of the American Metheorological Society, 95, 243–267, 2014.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M.: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Sci. Adv. 2020, 6, 2020.

Melton, J. R. and Arora, V. K.: Competition between plant functional types in the Canadian Terestrial Ecosystem Model (CTEM) v. 2.0, Geoscientific Model Development, 9, 323–361, 2016.

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R. S., and Fyfe, J. C.: The Canadian Seasonal to Interannual Prediction System. Part I: Models and Initialization, Monthly Weather Review, 141, 2910–2945, 2013.

Merryfield, W. J., Baehr, J., Batt, L., Becker, E. J., and Coauthors: Current and emerging developments in subseasonal to decadal prediction, Bulletin of the American Meteorological Society, https://doi.org/https://doi.org/10.1175/BAMS-D-19-0037.1, 2020.

Mohino, E., Rodriguez-Fonseca, B., Losada, T., Gervois, S., Janicot, S., Bader, J., Ruti, P., and Chauvin, F.: Changes in the interannual SST-forced signals on West African rainfall. AGCM intercomparison, Climate Dynamics, 37, 1707–1725, 2011a.

Mohino, E., Rodriguez-Fonseca, B., Mechoso, C. R., Gervois, S., Ruti, P., and Chauvin, F.: Impacts of the Tropical Pacific/Indian Oceans on the Seasonal Cycle of the West African Monsoon, Journal of Climate, 24, 3878–3891, 2011b.

Monerie, P.-A., Robson, J., Dong, B., and Dunstone, N.: A role of the Atlantic Ocean in predicting summer surface air temperature over North East Asia, Climate Dynamics, 51, 473–491, 2018.

Myhre, G., Forster, P. M., Samset, B. H., Hodnebrog, O., Sillmann, J., Aalbergsjo, S. G., Andrews, T., Boucher, O., Faluvegi, G., Flaschner, D., Iversen, T., Kasoar, M., Kharin, V. V., Kirkevag, A., Lamarque, J.-F., Olivie, D., Richardson, T. B., Shindell, D., Shine, K. P., Stjern, C. W., Takemura, T., Voulgarakis, A., and Zwiers, F.: PDRMIP: A Precipitation Driver and Response Model Intercomparison Project – Protocol and Preliminary Results, Bulletin of the American Metheorological Society, 98, 1185–1198, 2017.

Nobre, P., Marengo, J. A., Cavalcanti, I. F. A., Obregon, G., Barros, V., and Camilloni, I.: Seasonal-to-Decadal Predictability and Prediction of South American Climate, Journal of Climate, 19, 5988–6004, 2005.

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., Doney, S. C., Dunne, J. P., Dutay, J.-C., Graven, H., Griffies, S. M., John, J. G., Joos, F., Levin, I., Lindsay, K., Matear, R. J., McKinley, G. A., Mouchet, A., Oschlies, A., Romanou, A., Schlitzer, R., Tagliabue, A., Tanhua, T., and Yool, A.: Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP), Geoscientific Model Development, 10, 2169–2199, 2017.

Pauly, D. and Christensen, V.: Primary production required to sustain global fisheries, Nature, 374, 255–257, 1995.

Rowell, D. P.: The impact of Mediterranean SSTs on the Sahelian rainfall season, Journal of Climate, 16, 849–862, 2003.

Rowell, D. P., Folland, C. K., Maskell, K., and Ward, M. N.: Variability of summer rainfall over tropical north Africa (1906–92) Observations and modelling, Q. J. R. Meteorol. Soc., 121, 669–704, 1995.

Ruprich-Robert, Y., Msadek, R., Castruccio, F., Yeager, S., Delworth, T., and Danabasoglu, G.: Assessing the Climate Impacts of the Observed Atlantic Multidecadal Variability Using the GFDL CM2.1 and NCAR CESM1 Global Coupled Models, Journal of Climate, 30, 2785–2810, 2017a.

Ruprich-Robert, Y., Msadek, R., Castruccio, F., Yeager, S., Delworth, T., and Danabasoglu, G.: Assessing the climate impacts of the observed Atlantic multidecadal variability using the GFDL CM2.1 and NCAR CESM1 Global Coupled Models, Journal of Climate, 30, 2785–2809, 2017b.

Ruprich-Robert, Y., Delworth, T., Msadek, R., Castruccio, F., Yeager, S., and Danabasoglu, G.: Impacts of the Atlantic multidecadal variability on North American summer climate and heat waves, Journal of Climate, 31, 3679–3700, 2018.

Scaife, A. A. and Smith, S.: A signal-to-noise paradox in climate science, npj Climate and Atmospheric Science, 2018.

Schiemann, R., Luthi, D., Vidale, P. L., and Schar, C.: The precipitation climate of Central Asia – intercomparison of boservational and numerical data sources in a remote semiarid region, International Journal of Climatology, 28, 295–314, 2008.

Sheen, K. L., Smith, D. M., Dunstone, N. J., Eade, R., Rowell, D. P., and Vellinga, M.: Skilful prediction of Sahel summer rainfall on inter-annual and multi-year timescales, Nature Communications, 8, 1–12, 2017.

Sienz, F., Müller, W. A., and Pohlmann, H.: Ensemble size impact on the decadal predictive skill assessment, Meteorolgische Zeitschrift, 2016.

Sigman, D. M. and Hain, M. P.: The Biological Productivity of the Ocean, Nature Education, 3, 2012.

Smith, D. M., Cusack, S., Colman, A. W., Folland, C. K., Harris, G. R., and Murphy, J. M.: Improved surface temperature prediction for the coming decade from a Global Climate Model, Science, 317, 796–799, 2007.

Smith, D. M., Eade, R., Dunstone, N. J., Fereday, D., Murphy, J. M., Pohlmann, H., and Scaife, A. A.: Skilful climate model predictions of multi-year north Atlantic hurricane frequency, Nature Geosciences, 3, 846–849, 2010.

Smith, D. M., Scaife, A. A., and Kirtman, B. P.: What is the current state of scientific knowledge with regard to seasonal and decadal forecasting?, Envrionmental Research Letters, 7, 1–11, 2012.

Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müler, W. A., Pohlmann, H., Yeager, S., and Yang, X.: Robust skill of decadal climate predictions, npj Climate And Atmospheric Science, 13, 1–10, 2019.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.-P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W. J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P.-A., Muller, W. A., Nicoli, D., Ortega, P., Pankatz, K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X., and Zhang, L.: North Atlantic climate far more predictable than models imply, Nature, https://doi.org/https://doi.org/10.1038/s41586-020-2525-0, 2020.

Smith, T. M., Reynolds, R., Peterson, T., and Lawrimore, J.: Improvements NOAAs Historical Merged Land-Ocean Temp Analysis, Journal of Climate, 21, 2283–2296, 2008.

Sospedra-Alfonso, R. and Boer, G. J.: Assessing the impact of initialization on decadal prediction skill, Geophsical Researh Letters, 47, 2020.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppA-assim, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3556, 2019a.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppA-hindcast, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3557, 2019b.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppC-forecast-addAgung, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3570, 2019c.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppC-forecast-addElChichon, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3571, 2019d.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppC-forecast-addPinatubo, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3572, 2019e.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppC-forecast-noAgung, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3573, 2019f.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppC-forecast-noElChichon, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3574, 2019g.
Sospedra-Alfonso, R., Lee, W., Merryfield, W. J., Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 DCPP dcppC-forecast-noPinatubo, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3575, 2019h.
Strommen, K. and Palmer, T. N.: Signal and noise in regime systems: A hypothesis on the predictaility of the North Atlantic Oscillation, Quarterly Journal of the Royal Meteorological Society, 145, 147–163, 2019.
Sutton, R. T. and Dong, B.: Atlantic Ocean influence on a shift in European climate in the 1990s, Nature Geoscience, 5, 788–792, 2012.

Swart, N. C., Cole, J., Kharin, S.and Lazare, M. S. J., Gillett, N., Anstey, J., Arora, V., Christian, J., Hanna, S., Jiao, Y., Lee, W., Majaess, F., Saenko, O., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model (CanESM), v5.0.3., https://doi.org/10.5281/zenodo.3251114, 2019a.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geoscientific Model Development, https://doi.org/doi.org/10.5194/gmd-2019-177, 2019b.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3610, 2019c.

Taguchi, B., Schneider, N., Nonaka, M., and Sasaki, H.: Decadal Variability of Upper-Ocean Heat Content Associated with Meridional Shifts of Western Boundary Current Extensions in the North Pacific, Journal of Climate, 30, 6247–6264, 2017.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 92, 485–498, 2012.

Tietsche, S., Balmaseda, M., Zuo, H., Roberts, C., Mayer, M., and Ferranti, L.: The importance of North Atlantic Ocean transports for seasonal forecasts, Climate Dynamics, 55, 1995–2011, https://doi.org/doi.org/10.1007/s00382-020-05364-6, 2020.

Ting, M., Kushnir, Y., Seager, R., and Li, C.: Forced and Internal Twentieth-Century SST Trends in the North Atlantic, Journal of Climate, 22, 1469–1481, 2009.

Tippett, M. K., Barlow, M., and Lyon, B.: Statistical correction of Central Southwest Asian winter precipitation simulations, International Journal of Climatology, pp. 1421–1433, 2003.

Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, Journal of Geophysical Research Atmosphere, 119, 2864–2889, 2014.

Tivy, A., Howell, S. E. L., Alt, B., McCourt, S., Chagnon, R., Crocker, G., Carrieres, T., and Yackel, J. J.: Trends and variability in summer sea ice cover in the Canadian Arctic based on the Canadian Ice Service Digital Archive, 1960–2008 and 1968–2008., Journal of Geophysical Research, 116, 2011.

Uppala, S. M., KAllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., amd J.-F. Mahfouf, A. P. M., Morcrette, J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961–3012, 2005.

Verseghy, D. L.: The Canadian Land Surface Scheme (CLASS): Its history and future, Atmosphere-Ocean, 38, 1–13, 2000.

Villamayor, J., Ambrizzi, T., and Mohino, E.: Influence of decadal sea surface variability on northern Brazil rainfall in CMIP5 simulations, Climate Dynamics, 51, 563–579, 2018.

von Salzen, K., Scinocca, J. F., McFarlane, N. A., Li, J., Cole, J. N. S., Plummer, D., Verseghy, D., Reader, M. C., Ma, X., Lazare, M., and Solheim, L.: The Canadian Fourth Generation Atmospheric Global Climate Model (CanAM4). Part I: Representation of Physical Processes, Atmosphere-Ocean, 51, 4–125, 2013.

Ward, M. N.: Diagnosis and short-lead time prediction of summer rainfall in tropical North Africa at interannual and multidecadal timescales, Journal of Climate, 11, 3167–3191, 1998.

Wilks, D. S.: Resampling hypothesis tests for autocorrelated fields, Journal of Climate, 10, 65–82, 1997.

Xue, Y., Smith, T., and Reynolds, R.: Interdecadal Changes of 30-Yr SST Normals during 1871-2000, Journal of Climate, 16, 1601–1612, 2003.

Yeager, S. G. and Robson, J. I.: Recent progress in understanding and predicting Atlantic decadal climate variability, Current Climate Change Report, 3, 112–127, 2017.

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and Lovenduski, N. S.: Predicting near-term changes in the Earth system, Bulletin of the American Metheorological Society, pp. 1867–1886, 2018.

Zahariev, K., Christian, J. R., and Denman, K. L.: Preindustrial, historical, and fertilization simulations using a global ocean carbon model with new parameterizations of iron limitation, calcification, and N2 fixation, Progress in Oceanography, 77, 56–82, 2008.

Zhang, R. and Delworth, T.: Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes, Geophysical Research Letter, 33, 1–5, 2006.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., and Dong, J.: A global moderate resolution dataset of gross primary production of vegetation for 2000–2016, Scientific Data, 4, 2017.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis-analysis system for ocean and sea ice: a description of the system and assessment, Ocean Science, 15, 778–808, 2019.

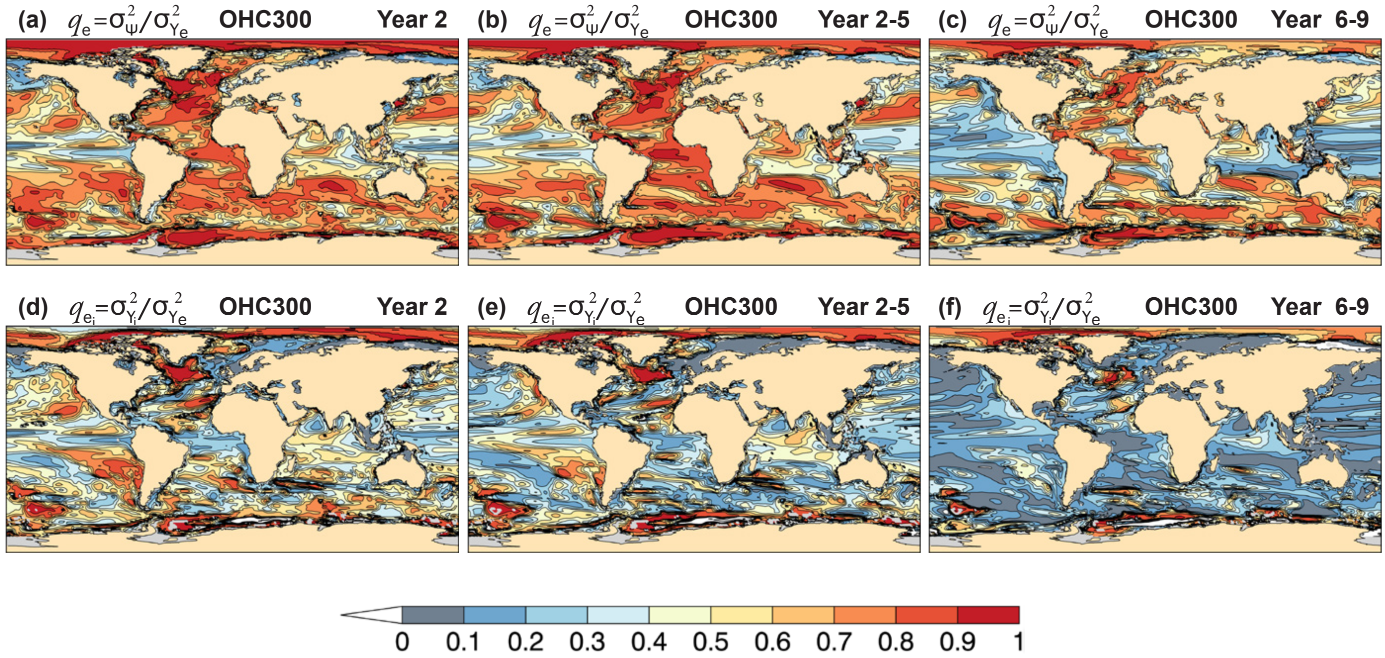

Figure 1. Potential predictability of CanESM5 ~~forecasts~~ hindcasts of annual and multi-year mean heat content above 300m. **(a-c)** Potentially predictable variance fraction $q_e = \sigma^2_\Psi/\sigma^2_{Y_e}$, Eq. (4), and **(d-f)** the portion $q_{e_i} = \sigma^2_{Y_i}/\sigma^2_{Y_e}$ attributed to initialization for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9.

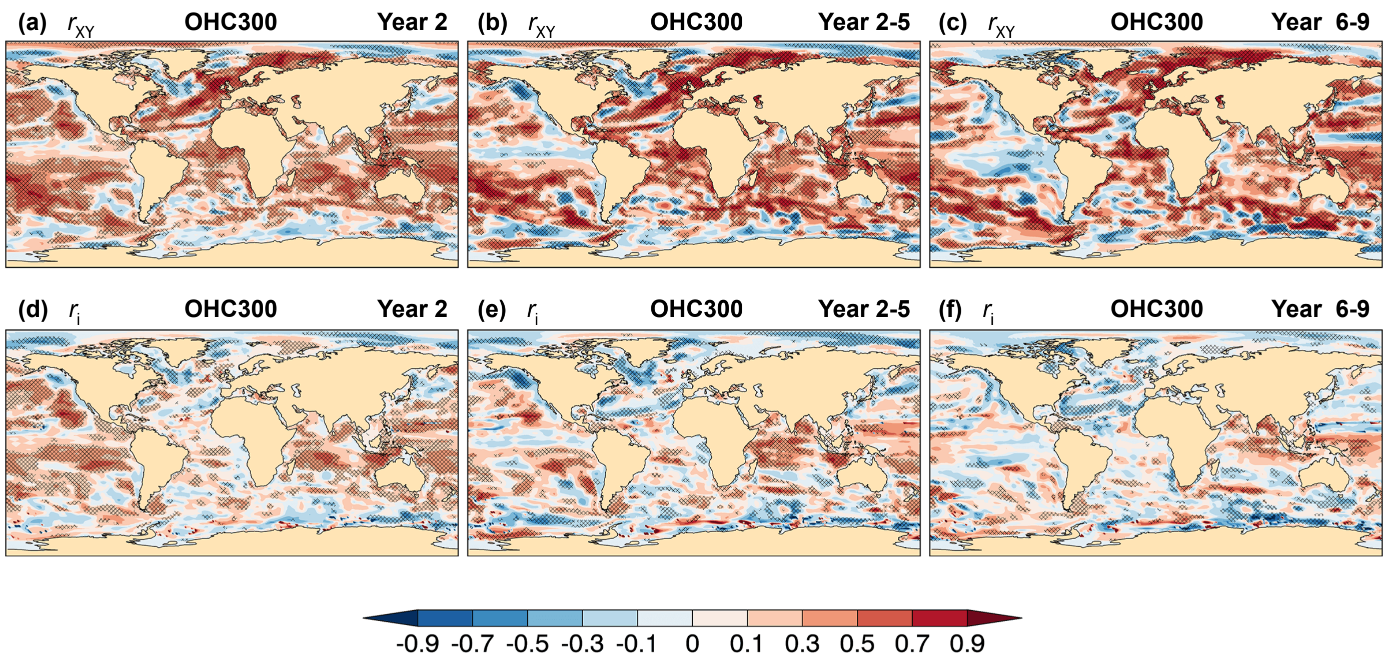

Figure 2. Skill of CanESM5 ~~forecasts~~ hindcasts of annual and multi-year mean heat content above 300m. **(a-c)** Correlation skill r_{XY}, Eq. (6), and **(d-f)** contribution from initialization r_i to correlation skill, Eq. (A17), for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9. The verifying observations are derived from EN4.2.1 dataset (appendix B). Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

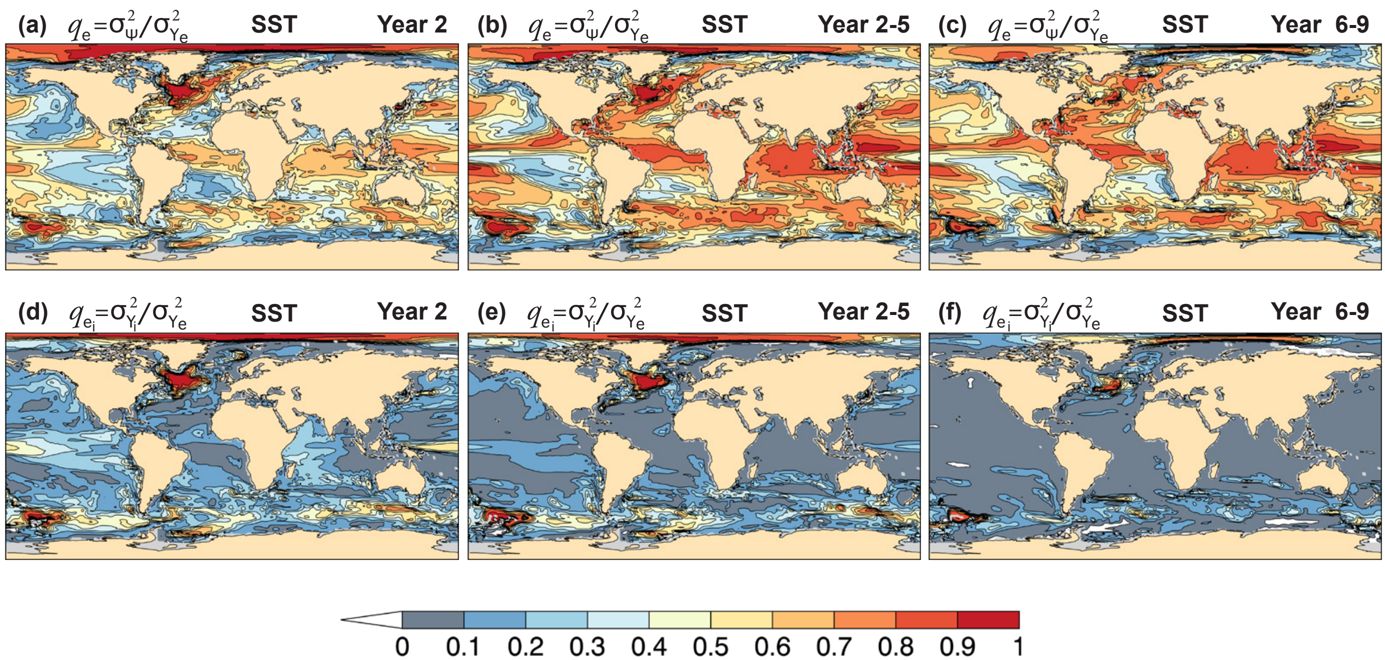

Figure 3. Potential predictability of CanESM5 annual mean sea surface temperature ~~forecasts~~ hindcasts. **(a-c)** Potentially predictable variance fraction $q_e = \sigma^2_\Psi/\sigma^2_{Y_e}$, Eq. (4), and **(d-f)** the portion $q_{e_i} = \sigma^2_{Y_i}/\sigma^2_{Y_e}$ attributed to initialization for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9.

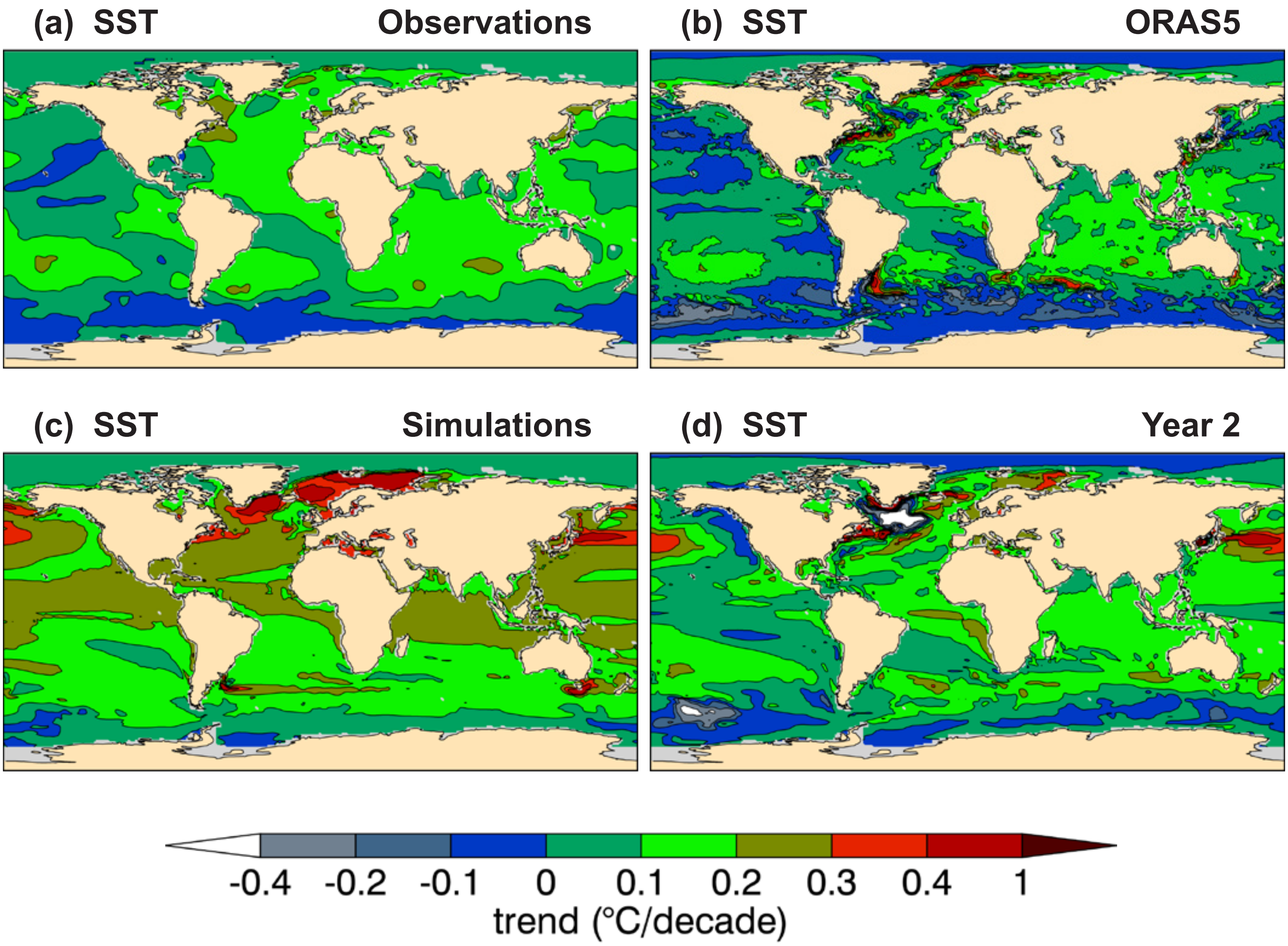

Figure 4. Linear trends of annual mean sea surface temperature for **(clockwise from upper-left)** observations, ORAS5, Year 2 ~~forecasts~~ hindcasts, and ~~historical~~ uninitialized simulations. The verifying observations are from ERSSTv5 dataset (appendix B).

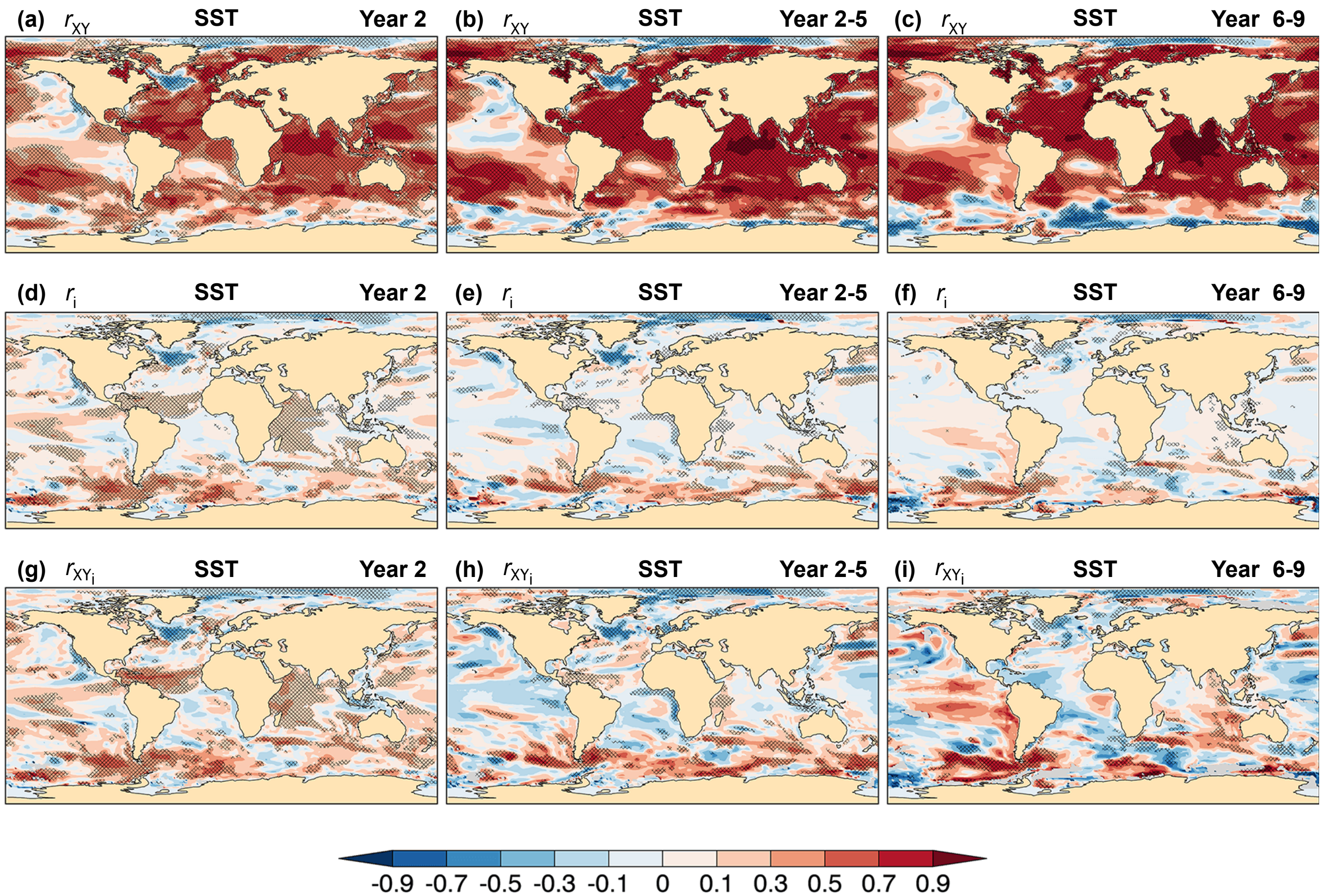

Figure 5. Skill of CanESM5 annual and multi-year mean sea surface temperature forecasts. **(a-c)** Correlation skill r_{XY}, Eq.(6), **(d-f)** contribution from initialization r_i to correlation skill, Eq. (A17), and **(g-i)** correlation skill of the initialized component of the forecast r_{XY_i}, Eq. (A19), for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9. The verifying observations are from ERSSTv5 dataset (appendix B). Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

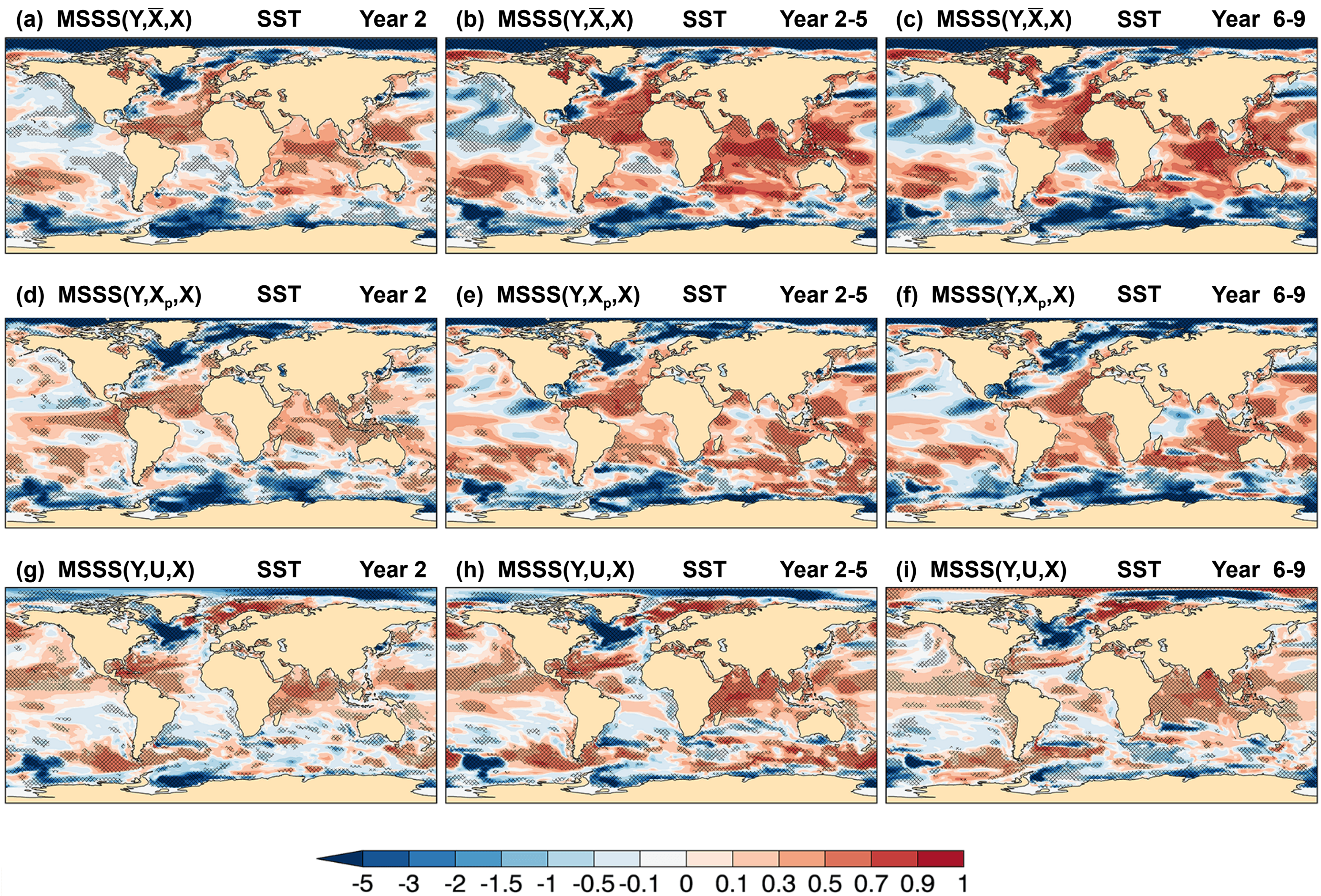

Figure 6. Skill of CanESM5 annual and multi-year mean sea surface temperature ~~forecasts~~ hindcasts. MSSS of **(left)** Year 2, **(middle)** Year 2-5 and **(right)** Year 6-9 ~~forecasts~~ hindcasts, Y, relative to **(a-c)** observed climatology, $\overline{X}$, **(d-f)** persistence forecast, X_p, and **(g-i)** ~~historical~~ uninitialized simulations, U. The verifying observations are from ERSSTv5 dataset (appendix B). Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

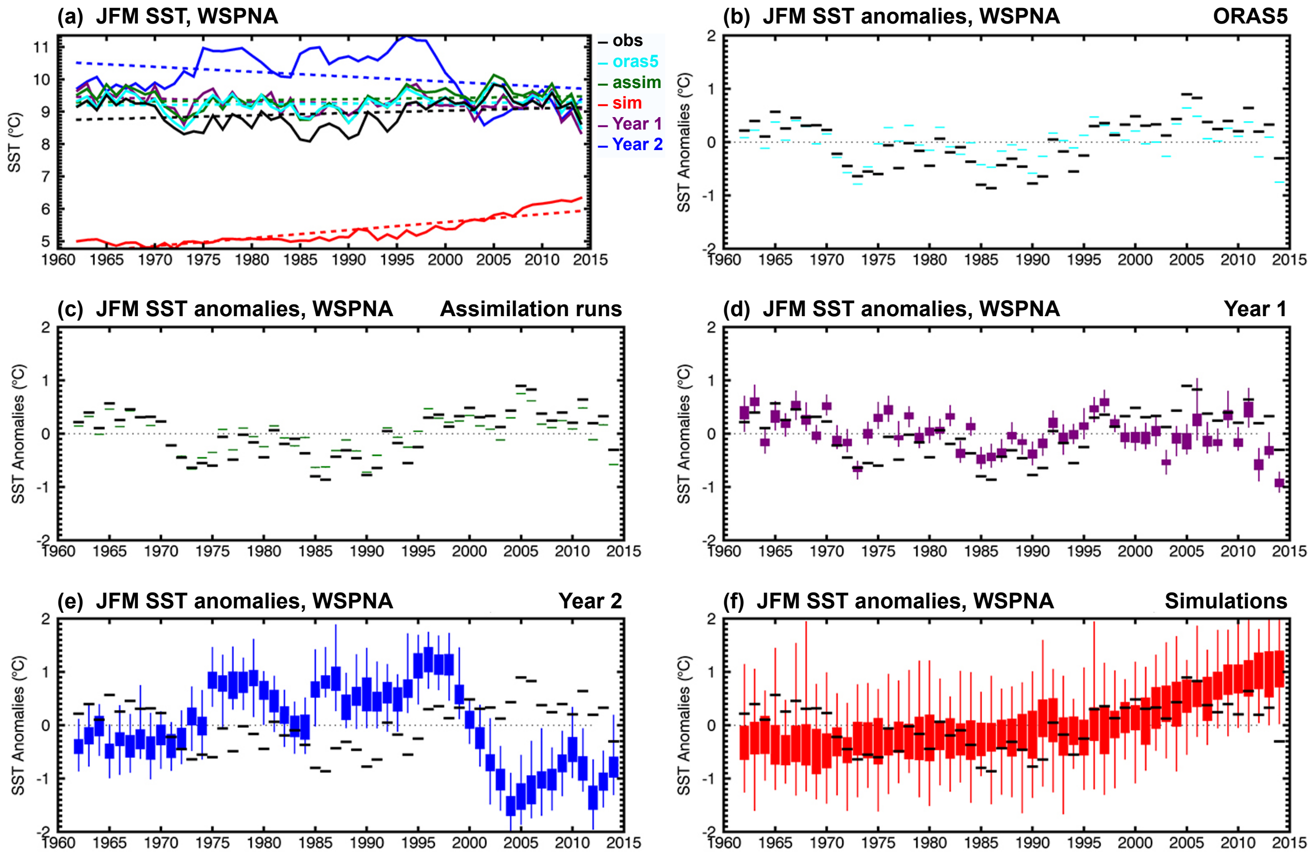

Figure 7. JFM time series of **(a)** SST and **(b-f)** SST anomalies corresponding to (black) ERSSTv5, (cyan) ORAS5, and CanESM5 (green) assimilation runs, (purple) Year 1 and (blue) Year 2 ~~forecasts~~ hindcasts, and (red) ~~historical~~ uninitialized simulations, averaged over the WSPNA region (40°N-60°N, 50°W-30°W). Boxes and whiskers indicate the minimum, maximum, 25- and 75-percentiles of the 40-member CanESM5 ensemble of forecasts and simulations, and the first 10-member ensemble for the assimilation runs. Model values in **(a)** correspond to ensemble means and dashed lines represent linear trends. Trends are not removed from the anomalies in **(b-f)**.

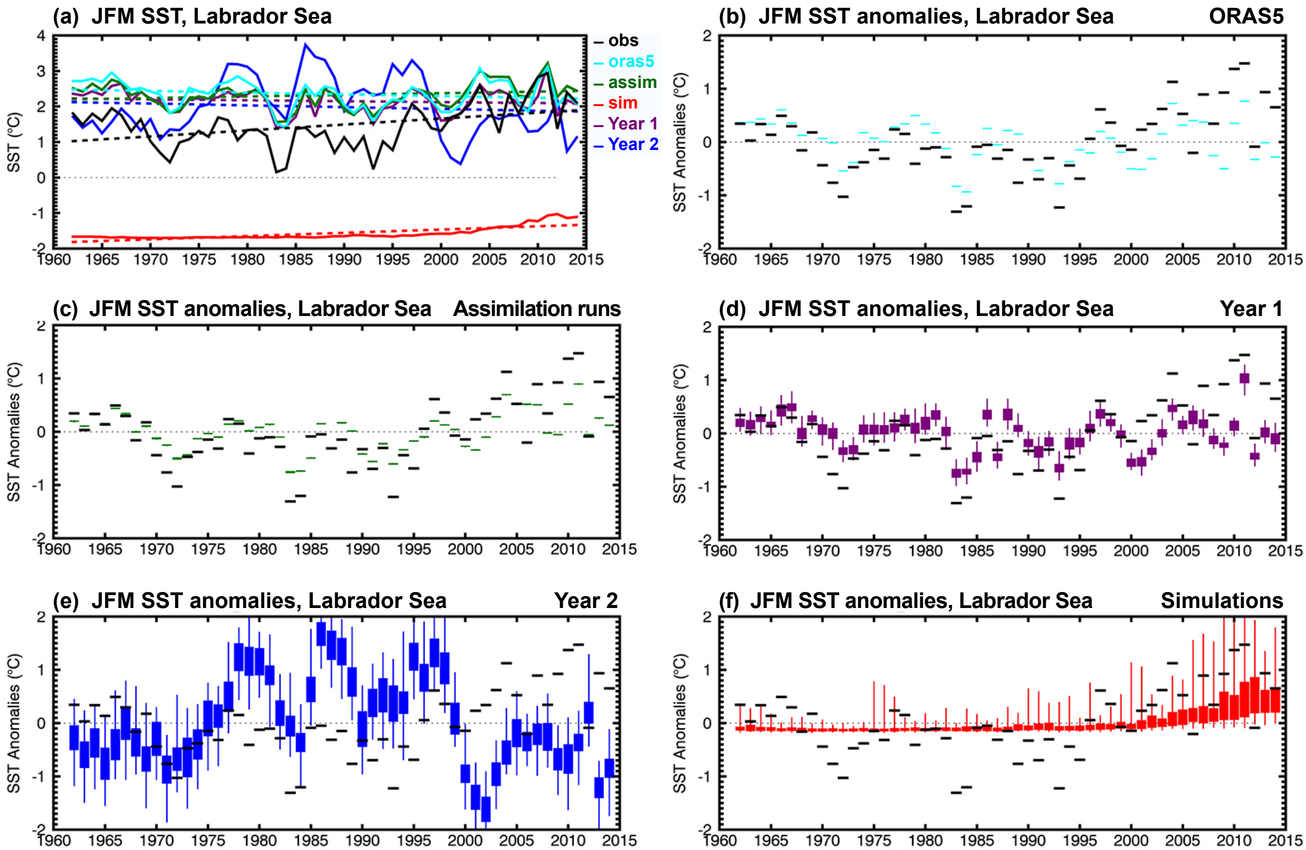

Figure 8. As in Fig. 7 for the Labrador Sea (55°N-65°N, 60°W-45°W).

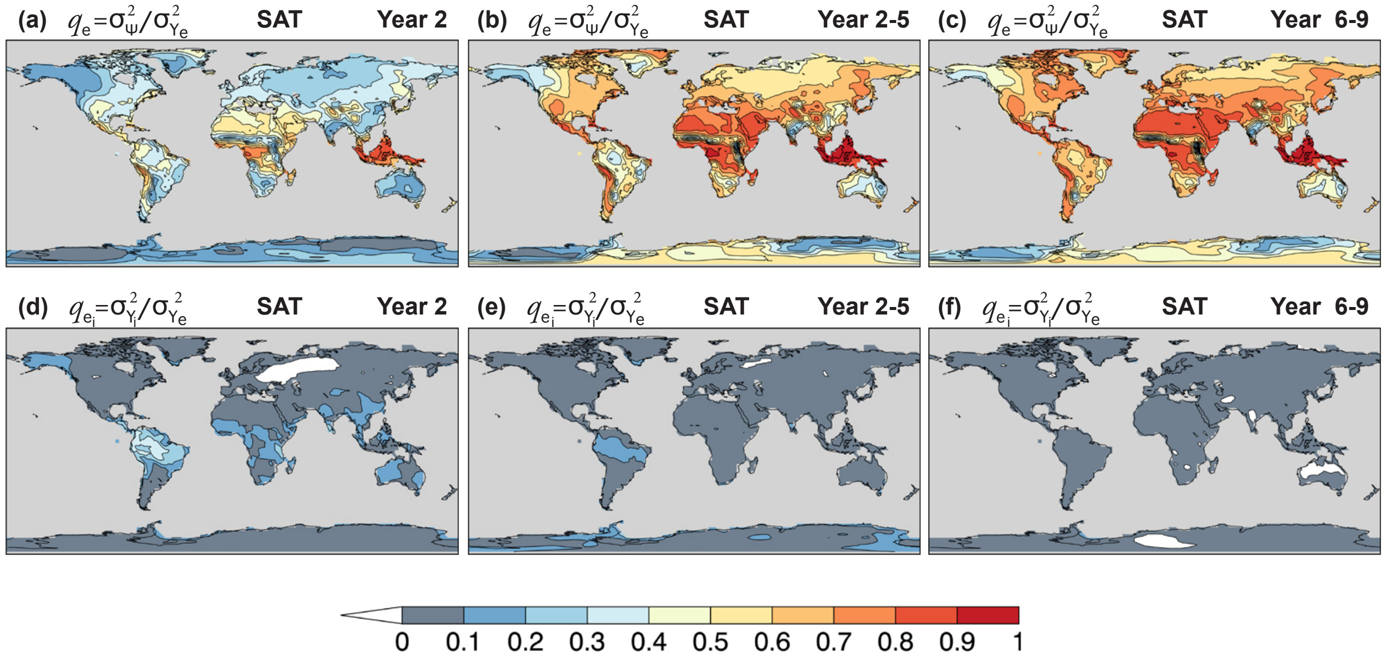

Figure 9. Potential predictability of CanESM5 ~~forecasts of~~ annual and multi-year mean near-surface air temperature ~~forecasts~~ hindcasts. **(a-c)** Potentially predictable variance fraction $q_e = \sigma^2_\Psi/\sigma^2_{Y_e}$, Eq. (4), and **(d-f)** the portion $q_{e_i} = \sigma^2_{Y_i}/\sigma^2_{Y_e}$ attributed to initialization for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9.

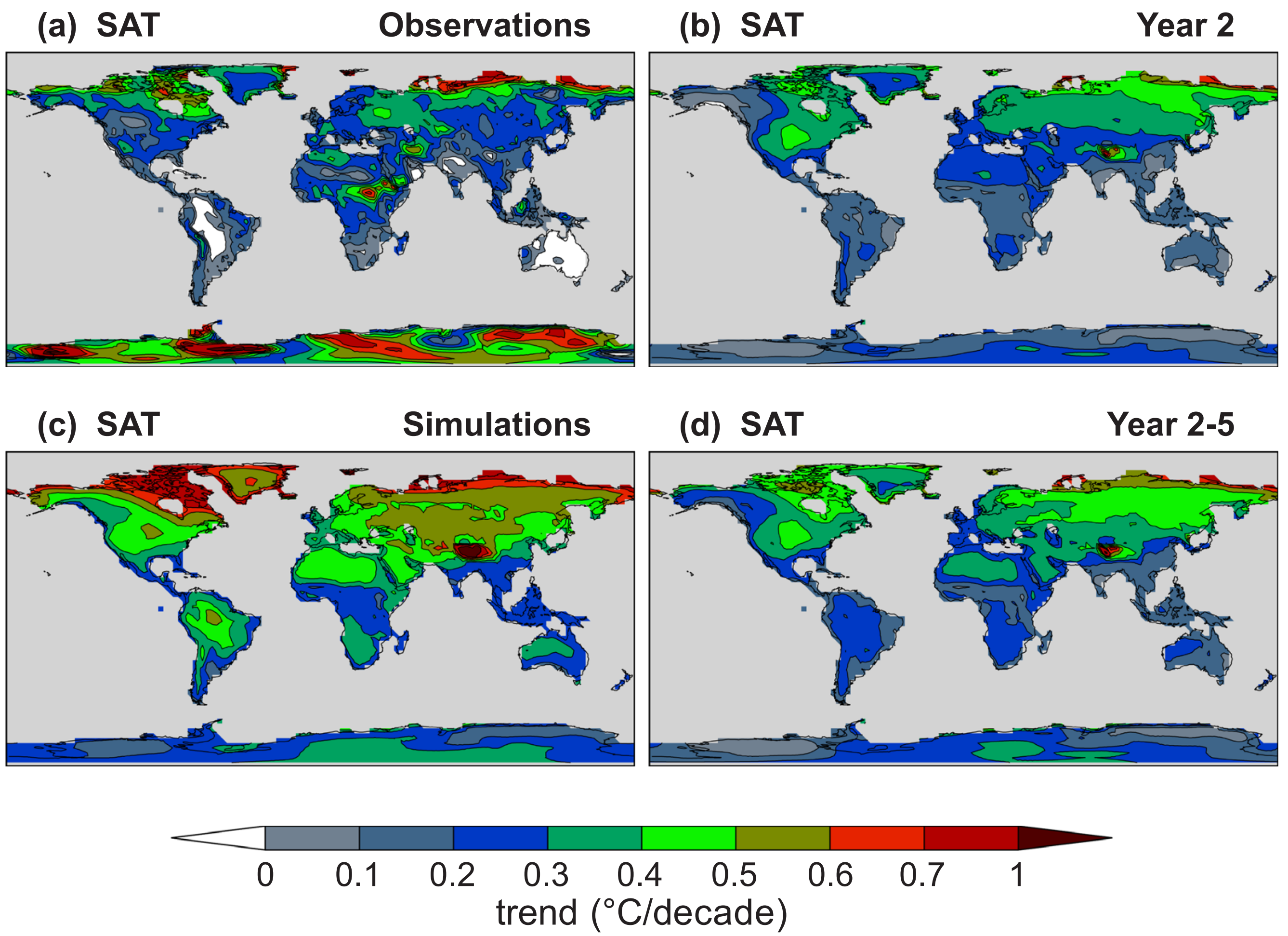

Figure 10. Decadal trends of annual mean near-surface air temperature for **(clockwise from upper-left)** observations, Year ~~1~~2 and Year 2-5 ~~forecasts~~ hindcasts, and ~~historical~~ uninitialized simulations. The verifying observations are from ERA-40 and ERA-Interim datasets (appendix B).

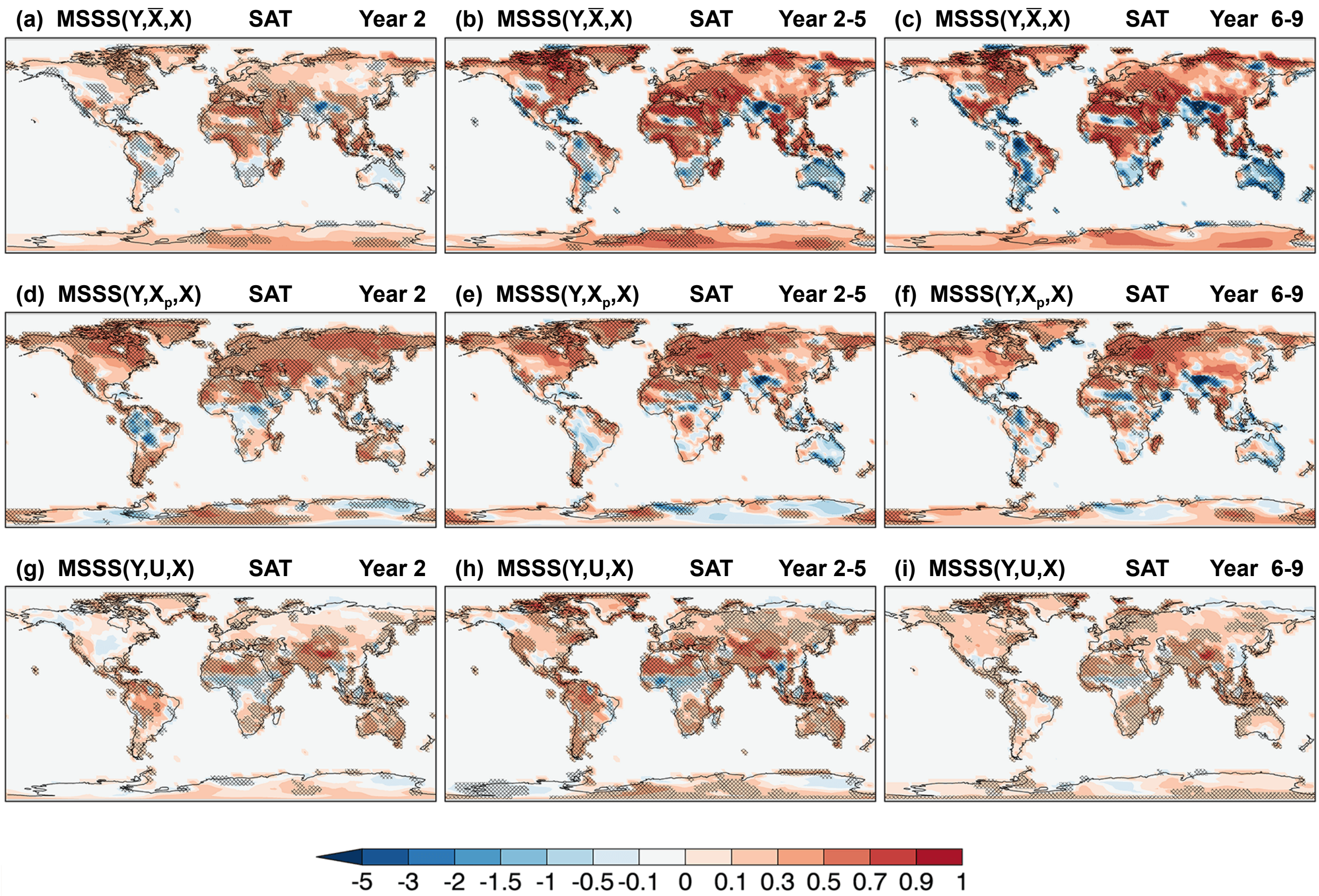

Figure 11. Skill of CanESM5 annual and multi-year mean near-surface air temperature ~~forecasts~~ hindcasts. MSSS of **(left)** Year 2, **(middle)** Year 2-5 and **(right)** Year 6-9 ~~forecasts~~ hindcasts, Y, relative to **(a-c)** observed climatology, $\overline{X}$, **(d-f)** persistence forecast, X_p, and **(g-i)** ~~historical~~ uninitialized simulations, U. The verifying observations are from ERA-40 and ERA-Interim datasets (appendix B). Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

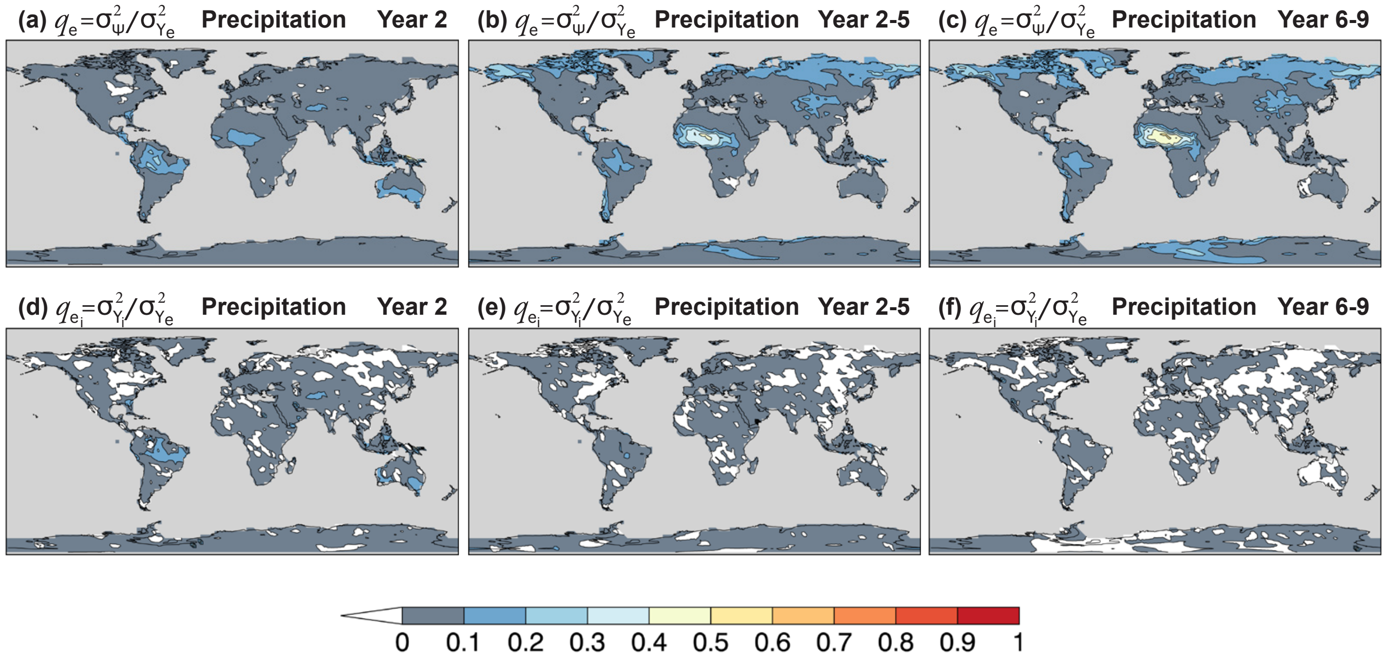

Figure 12. Potential predictability of CanESM5 ~~forecasts of~~ annual and multi-year precipitation hindcasts. **(a-c)** Potentially predictable variance fraction $q_e = \sigma^2_\Psi/\sigma^2_{Y_e}$, Eq. (4), and **(d-f)** the portion $q_{e_i} = \sigma^2_{Y_i}/\sigma^2_{Y_e}$ attributed to initialization for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9.

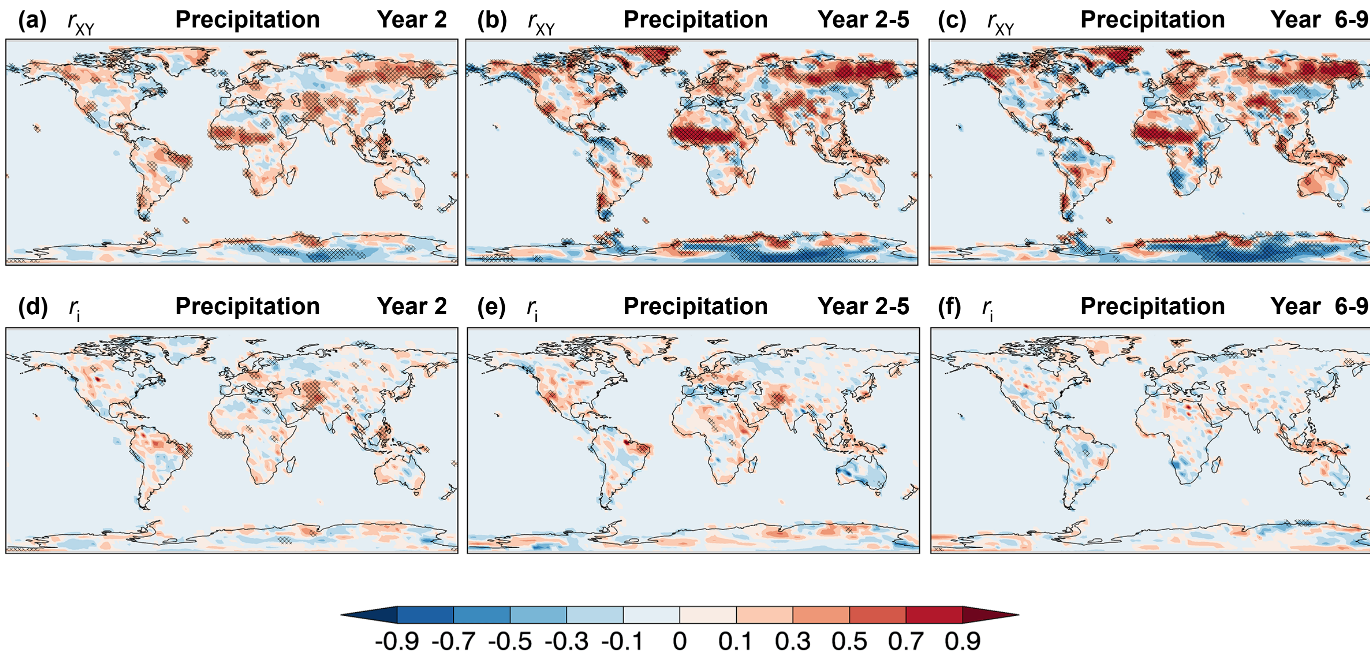

Figure 13. Skill of CanESM5 annual and multi-year mean precipitation ~~forecasts~~ hindcasts. **(a-c)** Correlation skill r_{XY}, Eq. (6), and **(d-f)** contribution from initialization r_i to correlation skill, Eq. (A17), for forecast **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9. The verifying observations are from GPCP2.3 dataset (appendix B). Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

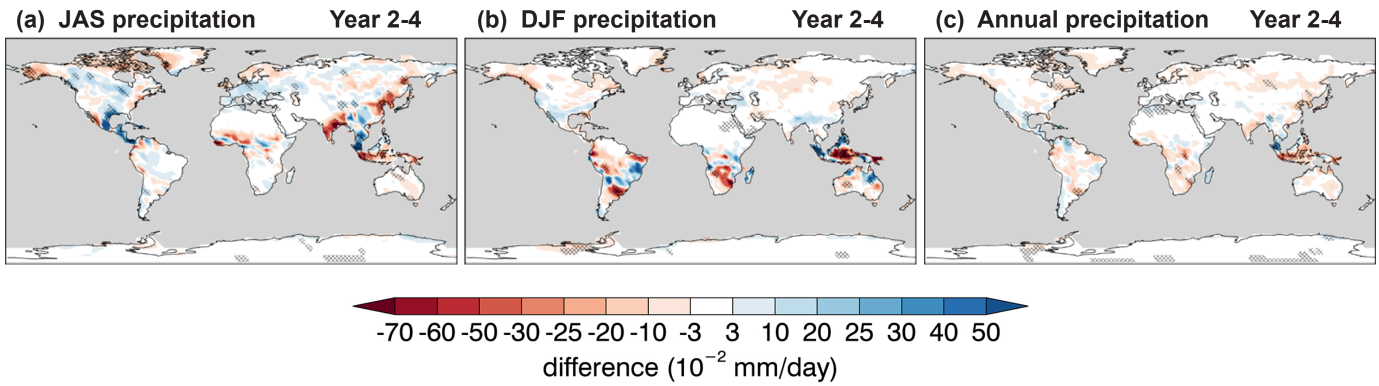

Figure 14. The impact of volcanic aerosols on mean precipitation forecasts. Difference of **(a)** summer (July-August-September), **(b)** winter (December-January-February), and **(c)** annual mean Year 2-4 precipitation forecasts with and without volcanic eruptions. Computations include 10-member ensembles of forecasts with and without eruptions of Mount Agung, El Chichón and Mount Pinatubo eruptions as per DCPP component C volcanic experiment setup (Boer et al., 2016). Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

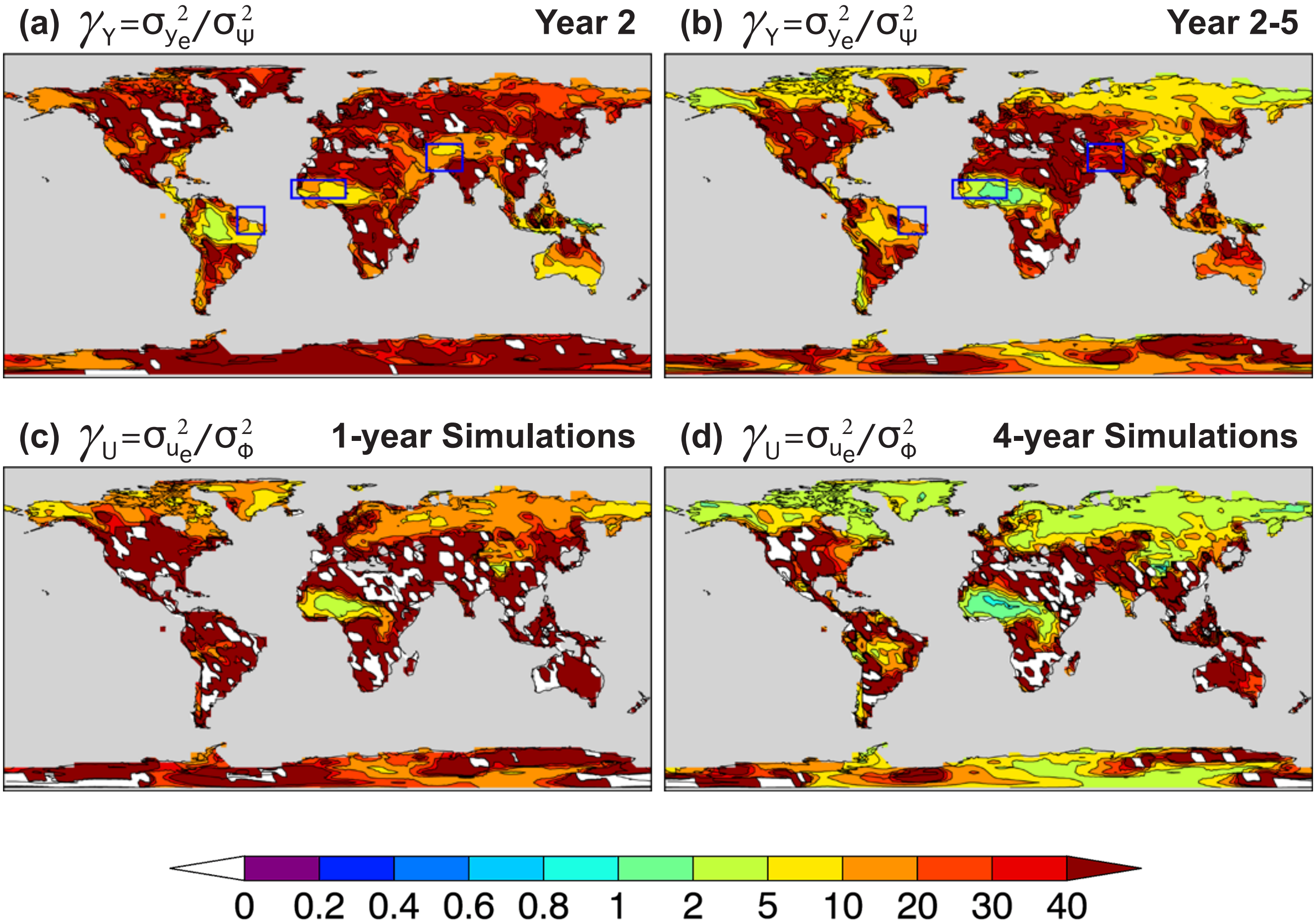

Figure 15. Maps of noise-to-predictable variance ratio **(a,b)** γ_Y, Eq. (A12), for **(a)** Year 2 and **(b)** Year 2-5 ~~forecasts~~ hindcasts and **(c,d)** γ_U, Eq. (A13), for **(c)** 1-year and **(d)** 4-year averaged uninitialized simulations of annual mean precipitation~~, produced with the 40-member ensembles of forecasts and simulations~~. The γ_Y ratio determines the ensemble size required to average out the noise component from the ensemble mean forecast, Eq. (5), and similarly γ_U for simulations. Negative values (white on land) result from sample errors, indicating small ensemble mean variance, and therefore small predictable signal, relative to the noise variance. The maps are produced with the full 40-member ensembles of hindcasts and uninitialized simulations. Rectangular boxes indicate the regions studied in Figs. 16-18 below.

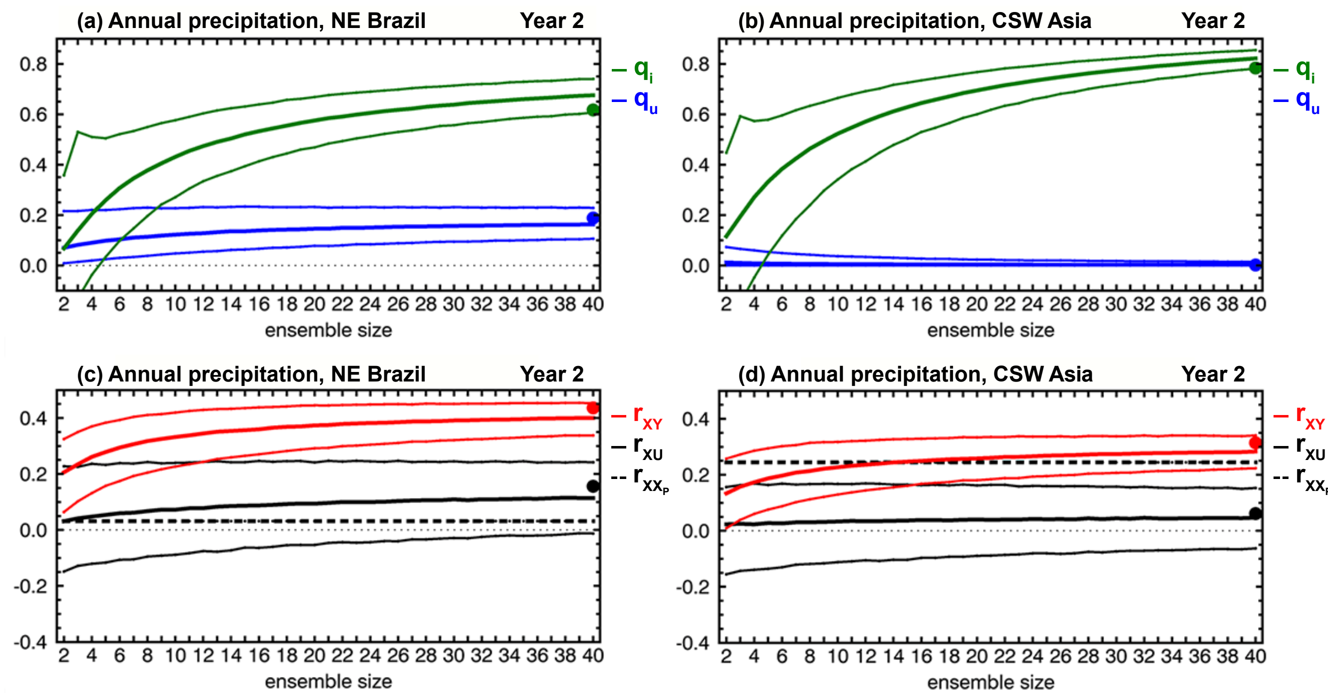

Figure 16. Dependence on ensemble size of **(a,b)** variance contributions $q_u = \sigma^2_{Y_u}/\sigma^2_Y$, Eq. (A14), in blue, and $q_i = \sigma^2_{Y_i}/\sigma^2_Y$, Eq. (A15), in green, to **(c,d)** correlation skill r_{XY}, Eq. (6), in red, of Year 2 ensemble mean precipitation ~~forecasts~~ hindcasts, averaged over **(a,c)** northeast Brazil (-10°N-5°N, 50°W-35°W) and **(b,d)** central Southwest Asia (25°N-55°N, 40°E-75°E). These regions are highlighted in Fig. 15 above. Thick black curves indicate correlation skill r_{XU} of ensemble mean uninitialized simulations. Thick dashed lines indicate correlation skill r_{XX_p} of the persistence forecast. Thin curves are confidence intervals derived from the 5th- and 95th-percentile of bootstrapping distributions generated from 10000 samples by random selection, with replacement, of ensemble members for each indicated ensemble size. Filled dots correspond to the actual 40-member ensemble predictions. Computations of q_u, Eq. (A14), and q_i, Eq. (A15), are done with $m_Y = 2 \ldots 40$ members from the ~~forecasts~~ hindcasts ensemble and, for each m_Y, the 40 members from the uninitialized simulations ensemble. The verifying observations used to compute correlation skill are from GPCP2.3 dataset (appendix B).

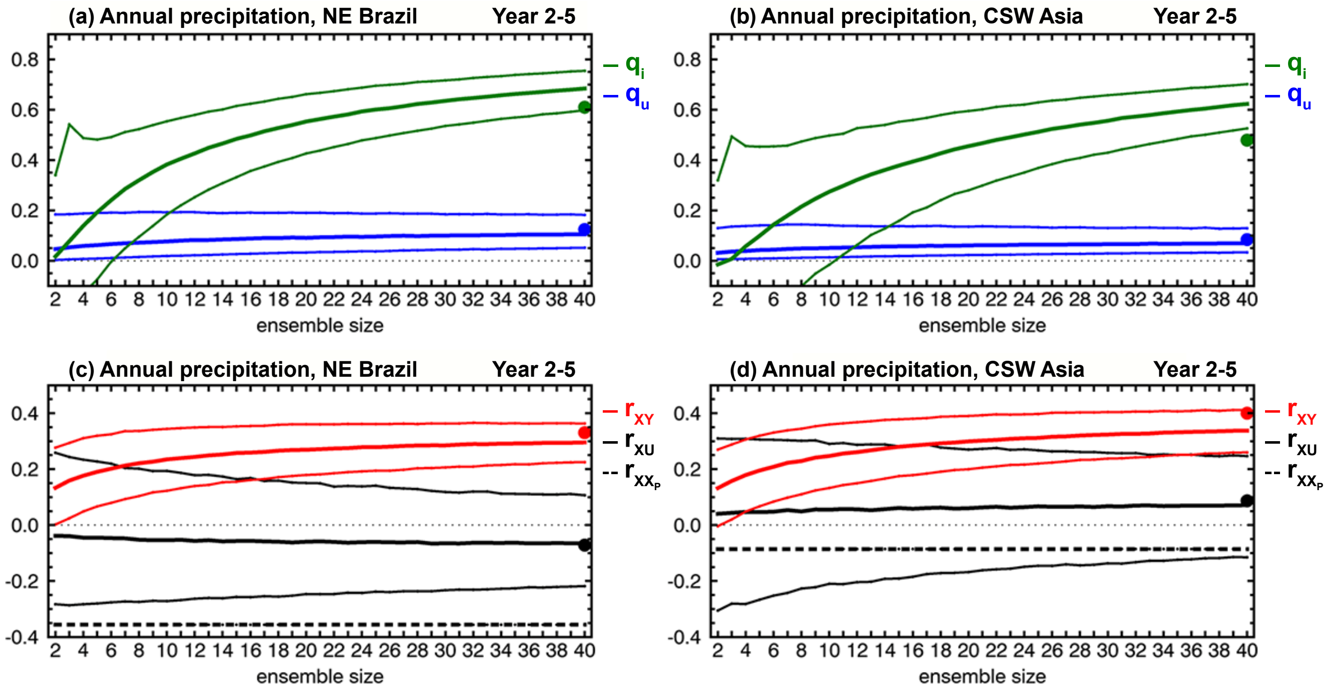

Figure 17. As in Fig. 16 for forecast year 2–5.

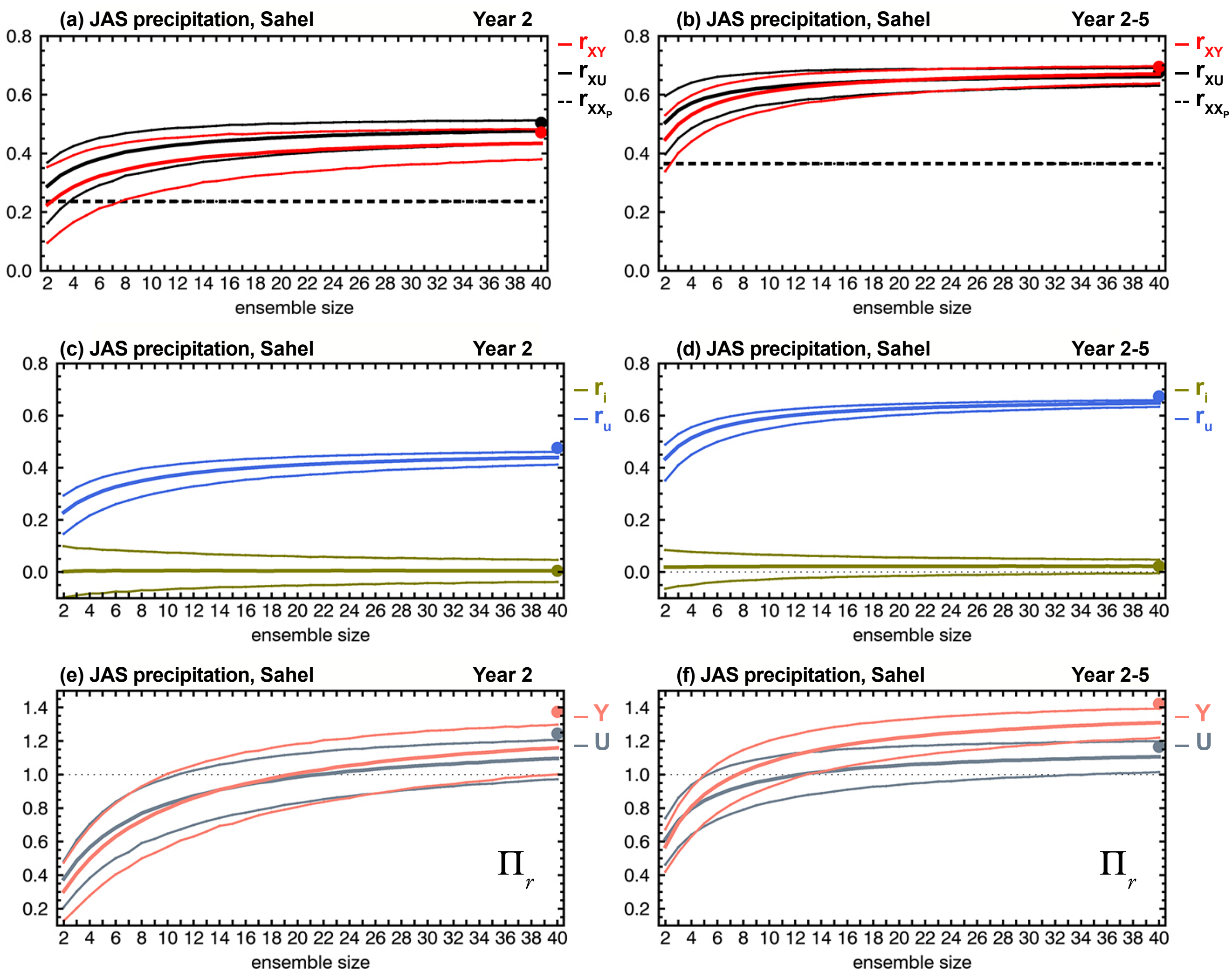

Figure 18. Dependence on ensemble size of **(a,b)** correlation skill of ensemble mean forecasts r_{XY} (red) and ensemble mean simulations r_{XU} (black); **(c,d)** contributions r_u, Eq. (A16), in royal blue, and r_i, Eq. (A17), in olive, to r_{XY}; and **(e,f)** ratio Π_r, Eq. (8), of ~~forecasts~~ hindcasts (salmon) and uninitialized simulations (gray), for **(a,c,e)** Year 2 and **(b,d,f)** Year 2-5 precipitation ~~forecasts~~ hindcasts, averaged over the Sahel (10°N-20°N, 20°W-10°E). This region is highlighted in Fig. 15 above. Thick dashed lines indicate correlation skill r_{XX_p} of the persistence forecast. Thin curves are confidence intervals derived from the 5th- and 95th-percentile of bootstrapping distributions generated from 10000 samples by random selection, with replacement, of ensemble members for each indicated ensemble size. Filled dots correspond to the actual 40-member ensemble predictions. Computations of r_u, Eq. (A16), and r_i, Eq. (A17), are done with $m_Y = 2\ldots40$ members from the ~~forecasts~~ hindcasts ensemble and, for each m_Y, the 40 members from the uninitialized simulations ensemble. The verifying observations used to compute correlation skill are from GPCP2.3 dataset (appendix B).

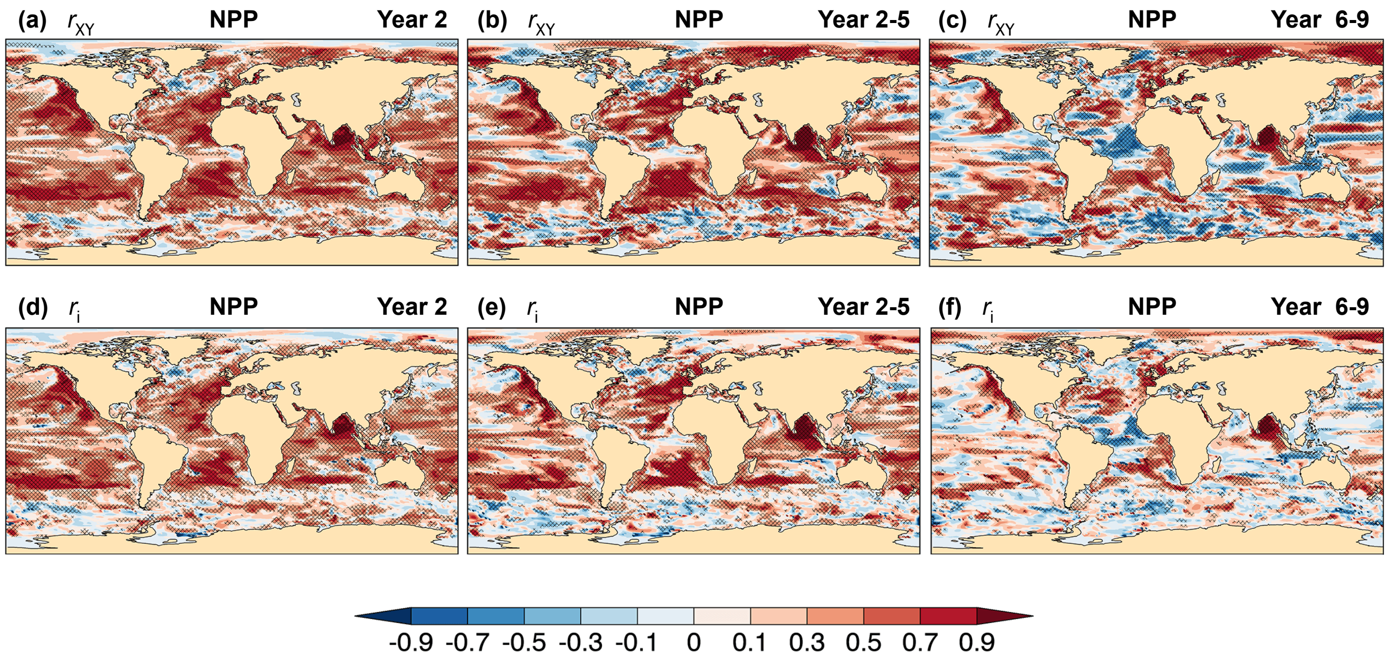

Figure 19. Potential for ~~S~~skill of CanESM5 annual and multi-year mean of ocean NPP ~~forecasts~~ hindcasts. **(a-c)** Correlation skill r_{XY}, Eq. (6), with the assimilation runs as verifying observations, and **(d-f)** contribution from initialization r_i to correlation skill, Eq. (A17), for **(left)** Year 2, **(center)** Year 2-5 and **(right)** Year 6-9 ~~forecasts~~ hindcasts. The CanESM5 assimilation runs used as verifying observations provide the initial conditions of the hindcasts. Cross-hatched regions indicate values significantly different from zero at the 90% confidence level.

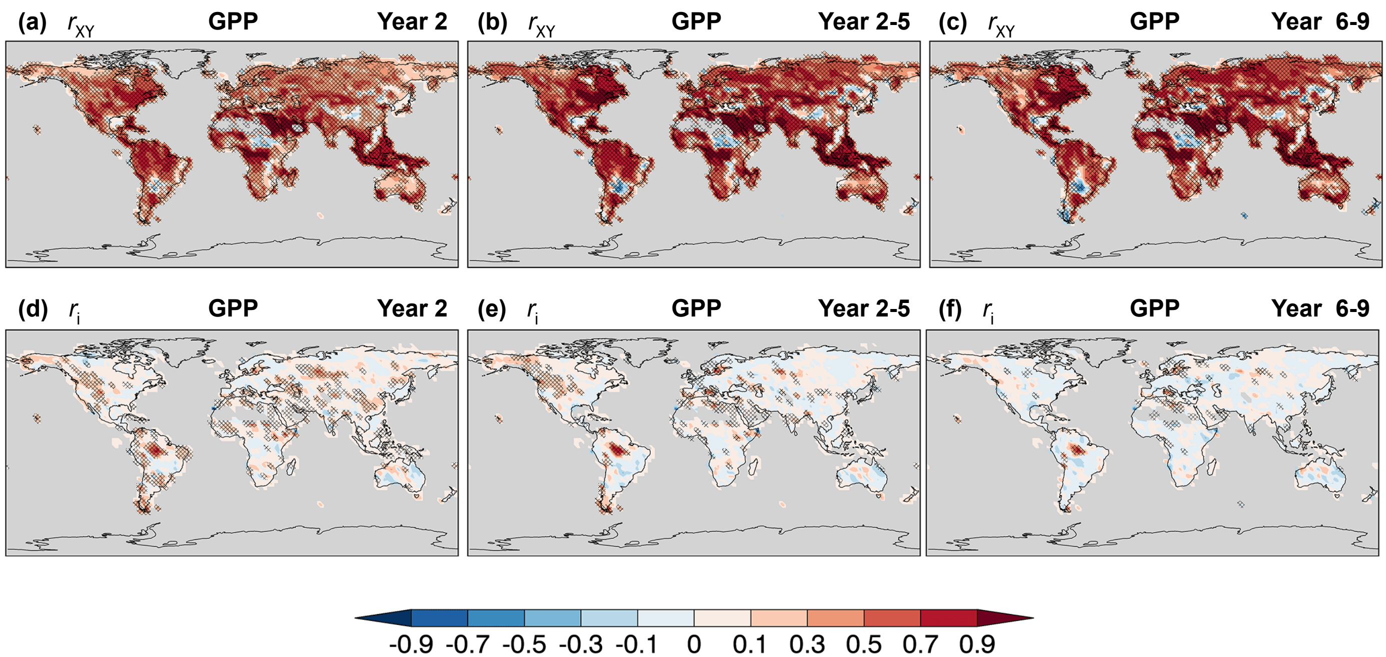

Figure 20. As in Fig. 19 for GPP on land.

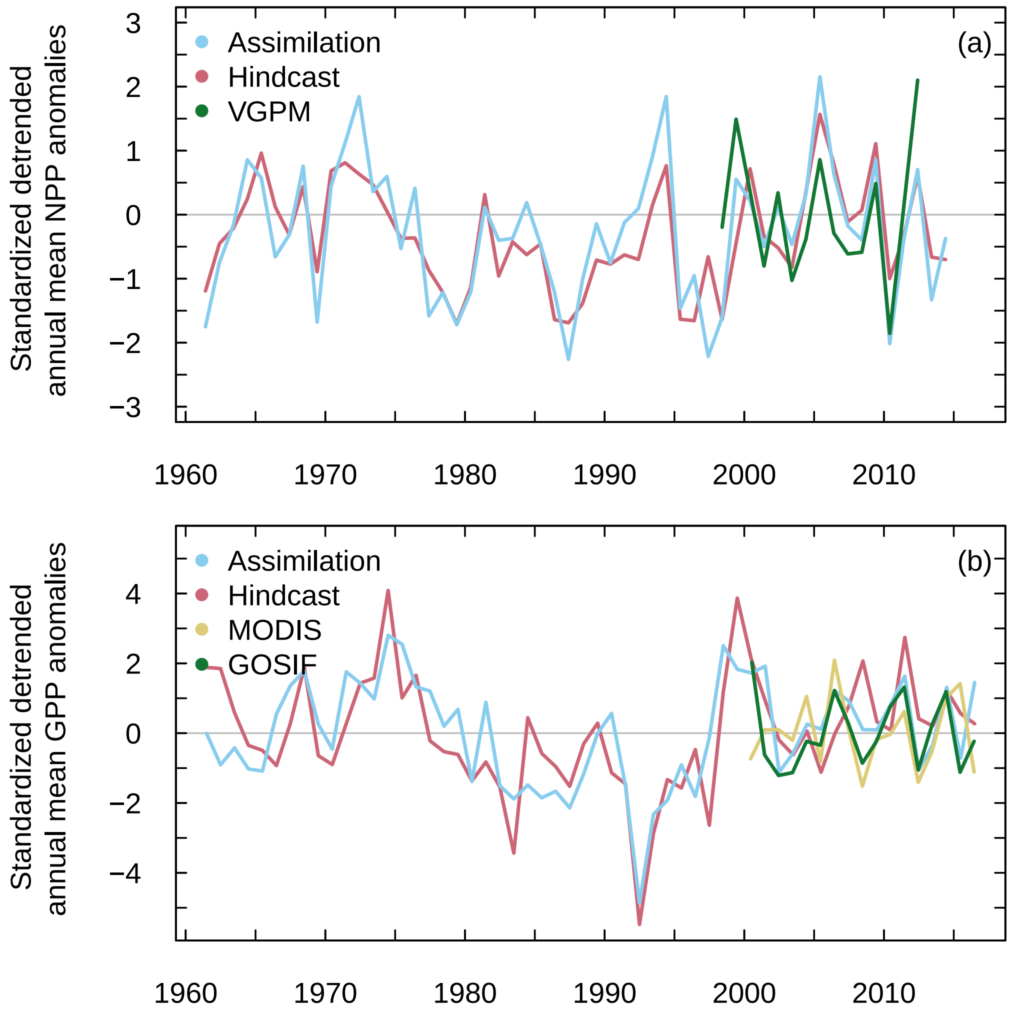

Figure 21. (a) Ocean integrated net primary productivity in the Canary Current region (10-18°W, 25-34°N), and **(b)** gross primary productivity on the global land, for the assimilation runs (blue) and Year 1 hindcasts (red). Observation-based estimates for **(a)** ocean, VGPM (green), and **(b)** land, MODIS (yellow) and GOSIF (green), are described in appendix B. Anomalies relative to the base period 2000-2016 have been linearly detrended and standardized.