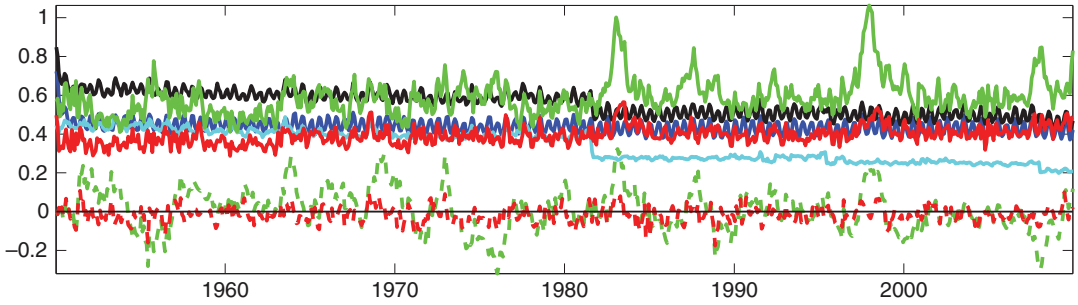

Fig. 1. Monthly RMSE (unit is in °C for all variables) of NorCPM (thick red line) and FREE (thick green line) against the assimilated SST anomaly. The contribution of the bias in the RMSE is plotted with the dashed line. The HadISST2 accuracy (i.e. the SD) is plotted in cyan. The ensemble spread of NorCPM is in blue and the black line is σ_{tot}, as defined in eq. (7).

monthly SST (Kennedy et al., 2015) and sea ice concentration (Titchner and Rayner, 2014) from 1850 to 2010 at 1° resolution. It is based on SST from ICOADS (International Comprehensive Ocean Atmosphere Data Set, http://icoads.noaa.gov/) and from the Met Office observational database; SST retrievals from AVHRR Pathfinder V5 data (1985–2007); and SST retrievals from the ATSR2 and AATSR (1995–2011) METEO products. The in-situ data are bias adjusted using an updated version of Kennedy et al. (2011). The ensemble samples the accuracy of the analysis by perturbing uncertain parameters. We use the ensemble spread to quantify the accuracy of the data set, which allows for time and space varying estimation of its accuracy. Furthermore, we assume that observation errors are decorrelated [i.e. **R** in eq. (6) is diagonal].

3. Global validation

Here, we assess the performance of NorCPM by comparing the monthly averaged re-analysis against the assimilated measurements (SST) and other independent measurements such as sea level, heat content (HC) and salt content (SC). NorCPM is compared to a corresponding 30-member ensemble integrated forward from the same initial conditions in 1950 and with the same external forcing as NorCPM, but without assimilation (hereafter referred as FREE). FREE would correspond to a typical climate projection exercise where the system is constrained only by the external forcing.

3.1. *Assimilated SST*

The performance of NorCPM in monitoring the variability of SST is assessed in terms of root mean square error (RMSE) and bias, and by comparing against FREE. NorCPM is expected to show less error than FREE, because the assessment is against the very same assimilated SST data. Nevertheless, the comparison is useful to assess the accuracy of the system over time. The statistics are computed using monthly averaged SST anomalies (w.r.t. the same 1950–2009 climatology used in the assimilation). The RMSE and the bias are calculated from the ensemble mean of NorCPM, FREE and observations. We also assess the reliability of NorCPM. The reliability is defined as the capability of the system to estimate its accuracy – the ensemble spread (the standard deviation of the ensemble) being used to quantify uncertainty.

DA reduces the RMSE and the bias consistently throughout the whole study period, with no obvious degradation (Fig. 1). There are pronounced maxima of RMSE in FREE corresponding to the large El Niño events (in 1982–1983, 1986–1987 and 1997–1998) during which the amplitude of the anomaly of SST is larger. Overall, the performance of FREE is poorer after 1982, which coincides with the availability of satellite data. The inclusion of this new data type in HadISST2 leads to a reduction of the observation error (cyan line) that is associated with better synchronisation among the observation members. From that time, the observation ensemble mean shows larger amplitude and smaller scale spatial structures, and it becomes less comparable to FREE. Comparatively, NorCPM shows only a slight increase in RMSE post 1982. Actually, NorCPM shows lower RMSE than the observational data set before 1982 but is slightly poorer afterwards.[1] In a perfect model framework, the error in an assimilation run would reduce with more accurate observations. However, the spatial scales of the features resolved in the observations post satellite era are smaller and their inherent predictability as well as the capability of our model to resolve them is reduced. Overall, the accuracy of NorCPM is stable with an accuracy of approximately 0.4 °C and no obvious bias.

[1]Note that the RMSE is calculated from the monthly averaged model outputs and not from the ensemble of model states at assimilation time (i.e. in the middle of the month). It is thus not ensured that the RMSE is lower than the observation error.