High predictive skill of global surface temperature a year ahead

C.K. Folland, A.W. Colman, D.M. Smith, Olivier Boucher, D.E. Parker, J.-P. Vernier

HAL Id: hal-01092299
https://hal.science/hal-01092299

Submitted on 8 Dec 2014

High predictive skill of global surface temperature a year ahead

Chris K. Folland,[1,2] Andrew W. Colman,[1] Doug M. Smith,[1] Olivier Boucher,[1,3]
David E. Parker,[1] and Jean-Paul Vernier[4,5]

[1]   We discuss 13 real-time forecasts of global annual-mean surface temperature issued by the United Kingdom Met Office for 1 year ahead for 2000–2012. These involve statistical, and since 2008, initialized dynamical forecasts using the Met Office DePreSys system. For the period when the statistical forecast system changed little, 2000–2010, issued forecasts had a high correlation of 0.74 with observations and a root mean square error of 0.07°C. However, the HadCRUT data sets against which issued forecasts were verified were biased slightly cold, especially from 2004, because of data gaps in the strongly warming Arctic. This observational cold bias was mainly responsible for a statistically significant warm bias in the 2000–2010 forecasts of 0.06°C. Climate forcing data sets used in the statistical method, and verification data, have recently been modified, increasing hindcast correlation skill to 0.80 with no significant bias. Dynamical hindcasts for 2000–2011 have a similar correlation skill of 0.78 and skillfully hindcast annual mean spatial global surface temperature patterns. Such skill indicates that we have a good understanding of the main factors influencing global mean surface temperature. **Citation:** Folland, C. K., A. W. Colman, D. M. Smith, O. Boucher, D. E. Parker, and J.-P. Vernier (2013), High predictive skill of global surface temperature a year ahead, *Geophys. Res. Lett.*, *40*, 761–767, doi:10.1002/grl.50169.

1.   Introduction

[2]   Global mean surface temperature (GST) varies with external forcings and modes of internal climate variability. The largest external forcings include those from anthropogenic greenhouse gases and aerosols, volcanic aerosols, and changes in solar output. The largest influencing mode of internal variability is the El Niño–Southern Oscillation (ENSO). Another internal predictor that varies slowly is the Atlantic Multidecadal Oscillation (AMO).

[3]   The motivation for the forecasts, which started in 2000, was first to gauge our understanding of the factors that influence GST on the interannual time scale. This has been particularly relevant as the forecasts have been made over a period of apparent reduced global warming which has caused considerable controversy [e.g., *Lawson*, 2008]. Second, the forecasts have provided useful information about the current state of the climate for international negotiations by governments on matters related to the United Nations Framework Convention on Climate Change at annual Conferences of the Parties (COP). Thus, some forecasts have been issued at the beginning of December to coincide with COP meetings and all forecasts have been issued in real-time Met Office press releases.

[4]   Forecasts up to 2010 were mainly created using a statistical approach. Although providing substantial real-time forecast skill, this system was not optimal. An updated version, still using almost the same factors that we believe influence GST, has been developed for forecasts from 2011. Since 2008, an ensemble of dynamical climate model forecasts starting from real initial conditions of the ocean and atmosphere has been included using the Met Office Decadal Prediction System [DePreSys, *Smith et al.*, 2007]. In principle, this includes similar influencing factors to the predictors used by the statistical methods including natural variations such as ENSO.

2.   Development of the 2000–2010 Forecasts

2.1.   Statistical Forecasts

[5]   The statistical forecast system was designed to predict the annual mean GST for the calendar year ahead using the contemporary Met Office GST data set. Table S1 shows the observed data sets used year by year to help train the statistical models. The most important GST data used here are HadCRUT3 [*Brohan et al.*, 2006], the latest version of the USA National Climate Data Center data (NCDC) [*Smith et al.*, 2008], and the GISTEMP data set from the Goddard Institute of Space Studies [*Hansen et al.*, 2010], here called GISS.

[6]   Two regression methods were created, both using in principle the same physical predictors. The two methods only differ in their representation of ENSO. In method 1, ENSO was represented statistically through eigenvectors (EOFs) of sea surface temperature (SST) described in *Folland et al.* [1999] and *Parker et al.* [2007] using training data usually commencing in 1947. In method 2 ENSO was predicted using a dynamical model. All predictors were chosen to reflect our best physical understanding of the interannual to interdecadal influences on GST. These included calculations of the net global mean forcing effect of the main anthropogenic radiative forcing factors [e.g., *Johns et al.*, 2003], ENSO, and estimates of volcanic and solar forcing and of the value of a 13 year smoothed AMO index in the year before the forecast. The latter is believed to have a small forcing effect

on GST [e.g., *Knight et al.*, 2005]. In both regression methods, the most important radiative forcing factor on decadal time scales, net anthropogenic radiative forcing, was the global mean net forcing due to greenhouse gases and anthropogenic aerosols (GA) observed or estimated for the previous year. This was taken from a time series of net radiative forcing calculated in HadCM3 transient climate change integrations [*Johns et al.*, 2003] (Supporting Information Text S1). Separate greenhouse gas and anthropogenic aerosol forcing predictors were not used in the statistical methods because of their substantial colinearity. Volcanic forcing data for the year prior to a forecast were updated from *Sato et al.* [1993] and solar forcing data updated from *Fröhlich and Lean* [1998] with the best available observations. These solar data are now known to have too much trend before 1950 [*Gray et al.*, 2010]; however, this caused no biases as training data only started in 1947 in method 1 and later than this in method 2. All the predictor data are available from the authors (Supporting Information S2 gives more detail).

[7] The cross-validated (Supporting Information Text S3) total correlation hindcast skill of the statistical models lay in the range 0.91–0.94, whereas the interannual hindcast correlation skill (using a high-pass filter, half power near 10 years) ranged from 0.67 to 0.74, corresponding to around 85% of total and about 50% of interannual variance. However, interannual skill estimates might be too high as insufficient weight was given to forcing from volcanic eruptions [compare equation (2) to equation (1) below]. Such volcanic forcing was particularly poorly known in the year prior to the first hindcast year after a major eruption, although several years with better prior data were always affected.

2.2. Example Statistical Forecast

[8] Equation (1) shows an example forecast for 2008 using method 1 trained against HadCRUT3 GST. The regression coefficients are derived from 1947 to 2006 standardized predictor training data.

$$G = 0.015 * A + 0.067 * E1 - 0.015 * E2 + 0.019 * V \quad (1)$$
$$+ 0.047 * S + 0.184 * GA + 0.027$$

where G = GST forecast anomaly from a 1961 to 1990 average (in °C), A = AMO index in 2007, $E1$ = ENSO EOF1 in October–November 2007, $E2$ = ENSO EOF2 in October to November 2007, S = solar forcing index in 2007, V = volcanic eruption forcing index in 2007, and GA = GA forcing in 2007.

[9] Equation (1) gives respective (rounded) contributions from the predictors and constant of 0.01, −0.08, −0.03, 0.02, −0.01, 0.41, and 0.03°C to a GST forecast of 0.36°C. As the observed GST using HadCRUT3 was 0.31°C, the error was 0.05°C, typical for method 1. Clearly, GA has most influence and ENSO EOF1 is the strongest interannual factor, especially as much of its variance is interannual. The solar factor had only a small influence on the forecast relative to observed GST in the previous year as it varies only slowly interannually. As discussed in section 5, the influence of the AMO index on GST was not optimally estimated. The volcanic eruption forcing weight was also likely too small but did not significantly affect these forecasts. The volcanic forcing influence was marginally positive in equation (1) because very little volcanic aerosol existed in the atmosphere in 2007 compared to the training period.

2.3. Dynamical Forecasts

[10] Dynamical coupled model forecasts from DePreSys [*Smith et al.*, 2007] were introduced in 2008. DePreSys is based on the third Hadley Centre climate model, HadCM3 [*Gordon et al.*, 2000], with a horizontal resolution of 2.5° latitude x 3.75° longitude in the atmosphere and 1.25° in the ocean. DePreSys includes the effects of changing concentrations of greenhouse gases and anthropogenic aerosols, projected changes in solar irradiance, and volcanic aerosol after known eruptions. By starting from the observed state of the ocean and atmosphere, DePreSys also has the potential to predict natural internal variability. This is achieved by relaxing HadCM3 to analyses of atmospheric winds, temperature and surface pressure from the European Centre for Medium Range Forecasts, and ocean temperature and salinity [*Smith and Murphy*, 2007]. Model drift is minimized by initializing with observed anomalies added to the model climatology. Forecasts consist of an ensemble of 10 members starting on 1st June and 10 members starting on 1st September. Although the forecasts extend for 10 years ahead, only the ensemble mean for the first full calendar year is used here. DePreSys was introduced because it had similar skill for GST forecasts 1 year ahead to the two statistical methods [*Smith et al.*, 2007].

3. Skill of Real-time Forecasts

[11] Figure 1 compares real-time GST forecasts for 2000–2011 with contemporary estimates of GST. For the latter, we have used the uncertainties estimated for HadCRUT3. Averages of the two statistical forecasts and the DePreSys predictions are also shown separately from 2008. Real-time forecast correlation skill is high at 0.75 (0.74 over 2000–2010) and root mean square error (rmse) correspondingly low at 0.07°C, with a mean warm bias of 0.05°C. The slightly greater warm bias of 0.06°C for the shorter period 2000–2010, relative to contemporary GST data, is significant at the 5% level; indeed, most forecasts appear slightly too
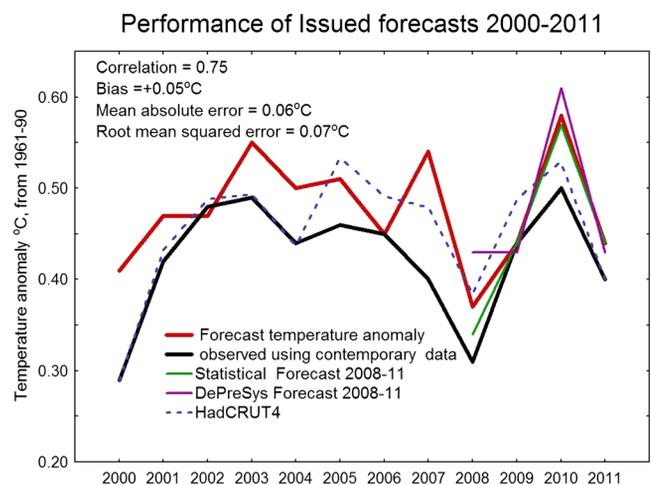
Figure 1. Performance of real-time issued forecasts, 2000–2011. Performance of issued forecasts relative to the contemporary version of HadCRUT. The separate statistical and dynamical forecast components since 2008 are shown. HadCRUT4 values (blue dashed) illustrate the cold bias in the original observations from 2004. Uncertainties are not shown for clarity.

warm. Despite this, all forecasts were within the 95% confidence limits of the observed GST. Errors greater than 0.1°C were confined to 2000 and 2007 and due mainly to forecasts of a stronger El Niño than actually occurred. The total correlation of 0.75 over 2000–2011 can be compared to a persistence total correlation of 0.37 and persistence rmse of 0.10°C. ENSO is the main interannual predictor, so a model based on ENSO alone and persistence of GST from the previous year has a much higher total correlation of 0.67 than persistence but an rmse of 0.12°C.

4. Warm Bias in 2000–2010 Forecasts

[12] The warm bias of 0.06°C in the 2000–2010 real-time forecasts has two identifiable causes. First, there is a recent cold bias in HadCRUT3 because of data gaps in the rapidly warming Arctic region [*Morice et al.*, 2012, Figure S2 and Supplementary Information Text S4]. Predecessors of HadCRUT3 are likely to have had broadly similar problems. Relative to HadCRUT4 GST, which contains much more recent Arctic data [*Morice et al.*, 2012], the cold bias in annual mean contemporary GST averages -0.04°C over 2000–2010, but much of the observed cold bias occurs from 2004 onward (Figure 1). A further component of the forecast warm bias, about 0.02–0.03°C, reflects use of inflated regression that erroneously inflates trend, although it also more appropriately increases interannual variance. The observed GST cold biases and inflated regression completely explain the apparent warm bias of 0.06°C over 2000–2010. However, HadCRUT4 may still be biased slightly cold as its coverage of the rapidly warming Arctic Ocean remains incomplete.

5. Improved Statistical Forecast Methodology and Performance

[13] After the 2010 forecast, significant changes were introduced in the statistical methodology and verification data. Statistical hindcasts using model ENSO are no longer used, so we describe a revised method 1R using a fixed predictand data set. We discuss the final version of these hindcasts; thus, real-time forecasts for 2011 and 2012 differ slightly. First, training data were extended back to 1891 to better represent the AMO forcing. Second, HadCRUT3 was replaced by the average of HadCRUT3, NCDC, and GISS GST data. The GISS analysis extrapolates up to 1200 km from the nearest station when stations are not available and so it estimates temperature over the whole Arctic region. Thus, GISS GST is generally warmest in the last few years, consistent with the very rapid warming of the central Arctic [*Simmons et al.*, 2010]. Consequently, the average of these three data sets is better than HadCRUT3 for estimating GST in recent training and verification data. At the time of this decision, HadCRUT4 was not available. Here, we summarize the main changes, and their impact.

[14] We updated GA data to be the global mean net radiative forcing derived from the Representative Concentration Pathway, RCP8.5, a data set being used by the Coupled Model Intercomparison project CMIP5. Decadal GA data are available from the RCP 2.0 database http://www.iiasa.ac.at/web-apps/tnt/RcpDb/dsd?Action=htmlpage&page=about (at November 27, 2012), which we have interpolated. RCP8.5 represents the combined radiative forcing of greenhouse gases (long-lived gases, tropospheric and stratospheric ozone) and anthropogenic aerosols (sulfate, black and organic carbon, biomass burning) [*Meinhausen et al.*, 2011]. This time series has been transformed using the *Held et al.* [2010] estimate of a 4 year short-term e-folding time response of GST to GA forcing. As RCP8.5 data are projected well into the future, we can use this response function for the year being forecast.

[15] The ENSO predictor was confined to ENSO EOF1. Because the October–November ENSO index used to make real-time predictions has a higher variance than that of ENSO training data based on a full year, we standardize this and all other predictors. The volcanic forcing predictor was revised using a new monthly series [*Vernier et al.*, 2011] that includes new information about recent minor volcanic forcing and revisions to forcing data in the satellite era. We introduced an e-folding time of 8 months to the volcanic forcing data to reflect the mean GST response to tropical volcanic forcing discussed in *Parker et al.* [1996]. In real-time forecasts and hindcasts, the estimated value of this e-folded forcing for the year prior to that forecast is used. Solar forcing was changed to an online data set due to Judith Lean providing a smaller multidecadal increase in solar forcing in the first half of the twentieth century (http://www.geo.fu-berlin.de/en/met/ag/strat/forschung/SOLARIS/Input_data/Calculations_of_Solar_Irradiance.pdf) (as at November 27, 2012), which we have updated from other sources. An e-folding time of 4 years is introduced to monthly solar data to reflect the GST response time to solar forcing, although this is uncertain [*Gray et al.*, 2010]. Again, in real-time forecasts and hindcasts, the estimated value of this e-folded forcing in the year prior to that forecast is used.

[16] The annual average AMO index of *Parker et al.* [2007] is now unsmoothed using its estimated value in the year prior to that forecast. Smoothing distorts sharp multiannual changes in the AMO that may occur such as in years centered near 1970 [*Thompson et al.*, 2010]. Annual detrended GST also lags annual unsmoothed AMO index by about 1 year over 1891–2011 with a lag 1 year correlation peak of 0.45, significant at the 1% level (Figure S3 and Supporting Information Text S5). All predictor data sets are available from the authors.

[17] Figure 2 shows time series of the new training data indices, e-folded where appropriate. These mostly differ relatively little from those used in the 2000–2010 forecasts for the overlapping training period, although using unsmoothed AMO data has some effect.

[18] Equation (2) shows the hindcast equation for 2008 based on training data for 1891–2006 standardized over 1891–1998 to compare with equation (1), noting that predictor weights change only slightly over 2000–2012.

[19]

$$G = 0.027 * A + 0.060 * E + 0.046 * V + 0.036 * S \qquad (2)$$
$$+0.146 * GA - 0.132$$

where A = AMO index in 2007, E = ENSO EOF1 in October to November 2007, S = e-folded solar forcing index in 2007, V = e-folded volcanic forcing index in 2007, and GA = RCP 8.5 e-folded GA forcing in 2008. Equation (2) gives (rounded) respective contributions from the predictors and constant of 0.01, -0.07, 0.02, 0.03, 0.49 and -0.13°C, with extra relative weight to the AMO and particularly volcanic forcing compared to equation (1). The solar effect is positive, despite the 2007 S index approaching the bottom of the solar cycle, because the training data includes many low S values
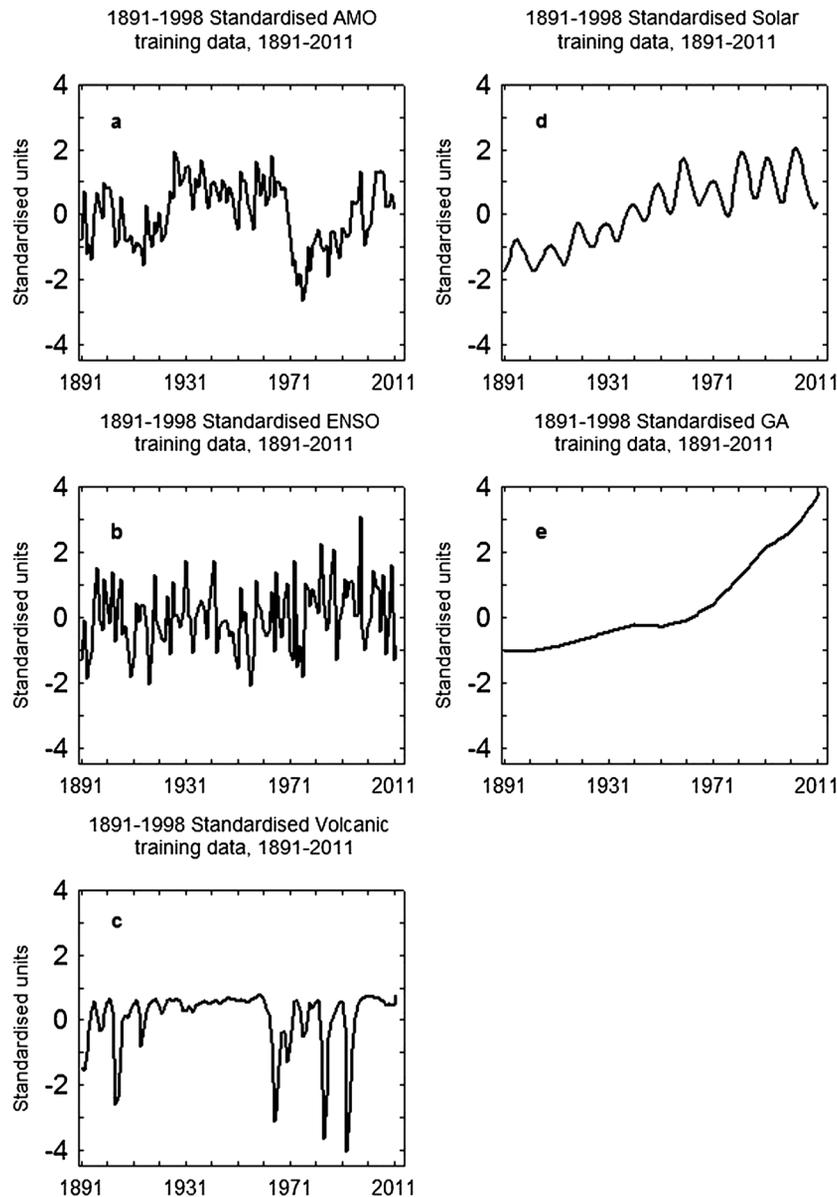
Figure 2. Time series of training data indices, 1891–2011, used in method 1R statistical hindcasts. Standardization is done for 1891–1998. Training data indices differ only slightly from those used for method 1 for the overlapping period from 1947, except that the AMO index is now unsmoothed. (a) AMO, (b) ENSO, (c) volcanic index, (d) solar index, (e) GA index. Standardization gives many small positive aerosol index values. The ENSO index is based on the high frequency SST EOF1 of *Parker et al.* [2007] and the AMO index is derived from the low frequency SST EOF3 of *Parker et al.* [2007].

before 1940. The hindcast of 0.35°C equals the observed value.

[20] Such equations allow tests of the potential impact of major tropical volcanoes that might not be initially accounted for. El Chichón gives an incremental forcing close to -3 standard deviations of V, giving an additional GST cooling of -0.13 or -0.14°C. A very large volcano like Pinatubo gives an incremental -4 standard deviations of V with an additional cooling near -0.18°C. These results are consistent with previous studies [e.g., *Parker et al.*, 1996]. Thus, the impact of major volcanoes whose forcing is not known at the time of the forecast can create a major forecast error. However, such forecasts can be updated once the additional volcanic forcing has been estimated.

[21] Figure 3 shows the performance of method 1R hindcasts for 2000–2011. Total correlation skill is 0.80, slightly higher than for issued forecasts over this period (0.75) with a similar rmse of 0.07°C. The bias of -0.03°C is not significant; clearly, there is no warm bias. For 2000–2011, we can compare method 1R, which continually updates the predictor equations over this period, with hindcasts where the predictor equations are fixed using training data up to 1998 to create the regression coefficients. These hindcasts tend to be slightly cooler than those created by method 1R, with a bias of -0.05°C, rmse of 0.08°C, but a similar total correlation skill of 0.81. Over the longer period 1951–2011, total cross-validated correlation skill is 0.95, although interannual skill is reduced to 0.55 compared to method 1. This period
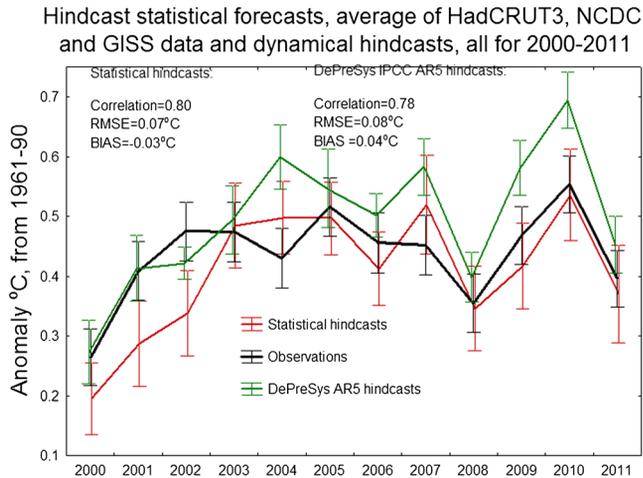
Figure 3. Performance of method 1R and DePreSys hindcasts, 2000–2011. Skill verified against the average of HadCRUT3, NCDC, and GISS data sets. Uncertainties for the observed data are set equal to those of HadCRUT3, those for DePreSys calculated from ensemble members. Uncertainties are ±1 standard deviation and for hindcasts are calculated for the better observed period from 1947; lines are offset slightly to allow uncertainties to be seen clearly.

includes errors in hindcasting several years with strong volcanic forcing, as the hindcast predictor for a given year is the value of V for the previous year. A simple hindcast method using ENSO and persistence alone has a lower total correlation of 0.86 over 1951–2011 and interannual correlation nearly disappears with a value of 0.10. Persistence alone gives negative interannual correlations (-0.22 for 1891–2011, -0.31 for 1951–2011). Removing just ENSO from the hindcast equations, total correlation is 0.94 but cross-validated interannual correlation reduces to 0.23 (0.17 over 1891–2011). The positive interannual correlations, both significant at the 10% level, indicate some interannual skill from the AMO, volcanic, and solar predictors (Figure S4, Table S2, and Supplementary Information Text S6), So these changes have removed the (mainly artificial) warm bias in the forecasts without compromising skill. DePreSys hindcasts prepared for CMIP5 for 2000–2011 (Figure 3) give quite similar results, except for a tendency to a small overall warm bias of 0.04°C. Their correlation with observations is 0.78 with an rmse of 0.08°C.

[22] Method 1R and DePreSys were used in the 2012 forecast with half weight each. The 2011 forecast differed by not including e-folding modifications to forcing data and the 2000–2010 GA forcing data were used. The 2011 forecast maintained the high skill and general similarity of DePreSys and statistical forecasts, with a statistical forecast of 0.44°C, a DePreSys forecast of 0.43°C, and an observed GST anomaly of 0.40°C. The 2012 forecasts are for 0.44°C (statistical method) and 0.52°C (DePreSys); provisional data indicate an observed value of 0.43°C.

6. Further Information from Dynamical Forecasts

[23] Figure 4a shows example DePreSys hindcasts for 2000–2010, which clearly indicate a potential to provide useful information on the spatial distribution of temperature.

Bracketed numbers are spatial correlation coefficients between each hindcast and the average of HadCRUT3, GISS, and NCDC data. Spatial skill is clearly considerably less than the GST skill in Figure 3, but many hindcasts pick up the ENSO signal quite well, giving an average spatial correlation of 0.42 for all years 2000–2011. Thus, DePreSys forecasts of surface temperature anomalies over some smaller areas than the globe are likely to be skillful. Figure 4b confirms this for an extended set of 50 hindcasts over 1960–2009 where most of the globe has significant correlations. Figure 4c shows that skill derived from initial conditions alone is significant worldwide, particularly over the oceans. The pattern of skill in Figure 4c indicates that its major component derives from ENSO, consistent with the statistical models. However, it is also clear from Figure 4b that changing radiative forcing over this 50 year period has been essential for skill as well, much of this forcing being anthropogenic (Figure 2). So it is not surprising that statistical and dynamical hindcasts 1 year ahead have similar skill over 2000–2011, as much of their skill derives from similar factors.

[24] Returning to the real-time forecasts and using HadCRUT4 data as being most reliable for hemispheric values, the correlation between DePreSys hindcast and observed hemispheric mean surface temperature anomalies over 2000–2011 was 0.81 (Northern Hemisphere) and 0.68 (Southern Hemisphere) with rmse of 0.08°C and 0.11°C, respectively. There is an insignificant warm bias of 0.02°C in the Northern Hemisphere but a larger warm bias of 0.08°C (just significant at the 5% level) in the Southern Hemisphere. The Southern Hemisphere appears to contain the main cause of a recent warm bias in the global DePreSys hindcasts in Figure 3.

7. Conclusions

[25] Eleven real-time GST forecasts issued by the Met Office over 2000–2010 explained about 55% of the largely interannual variance with a small but significant apparent warm bias of 0.06°C. The majority of this bias is likely to be due to cold biases in the verification GST data, particularly from 2004 to 2010, arising from data gaps in the rapidly warming Arctic. This was confirmed by higher values of GST in HadCRUT4 from 2004. The average of HadCRUT3, GISS, and NCDC data currently used is much less biased cold.

[26] Our ability to make skillful GST forecasts a year ahead during a period of slow warming indicates that we have a good understanding of factors influencing current and recent GST. The statistical method is mainly sensitive to the time-varying profiles of forcings, including GA forcing, so the absolute value of GA forcing could be quite poorly known due to substantial uncertainties in anthropogenic aerosol forcing yet still yield very good forecast skill. DePreSys GST forecasts are more sensitive to absolute GA forcing, although its influence is constrained a year ahead by the use of observed initial ocean conditions. Nevertheless, skillful and similar statistical and dynamical hindcasts over 2000–2011 suggest that this potential problem is not serious.

[27] We conclude that skillful GST forecasts 1 year ahead can be made in most years by both statistical and dynamical methods. Skill is mainly limited by our ability to forecast ENSO accurately and occasionally by cooling from unexpected large volcanic eruptions.
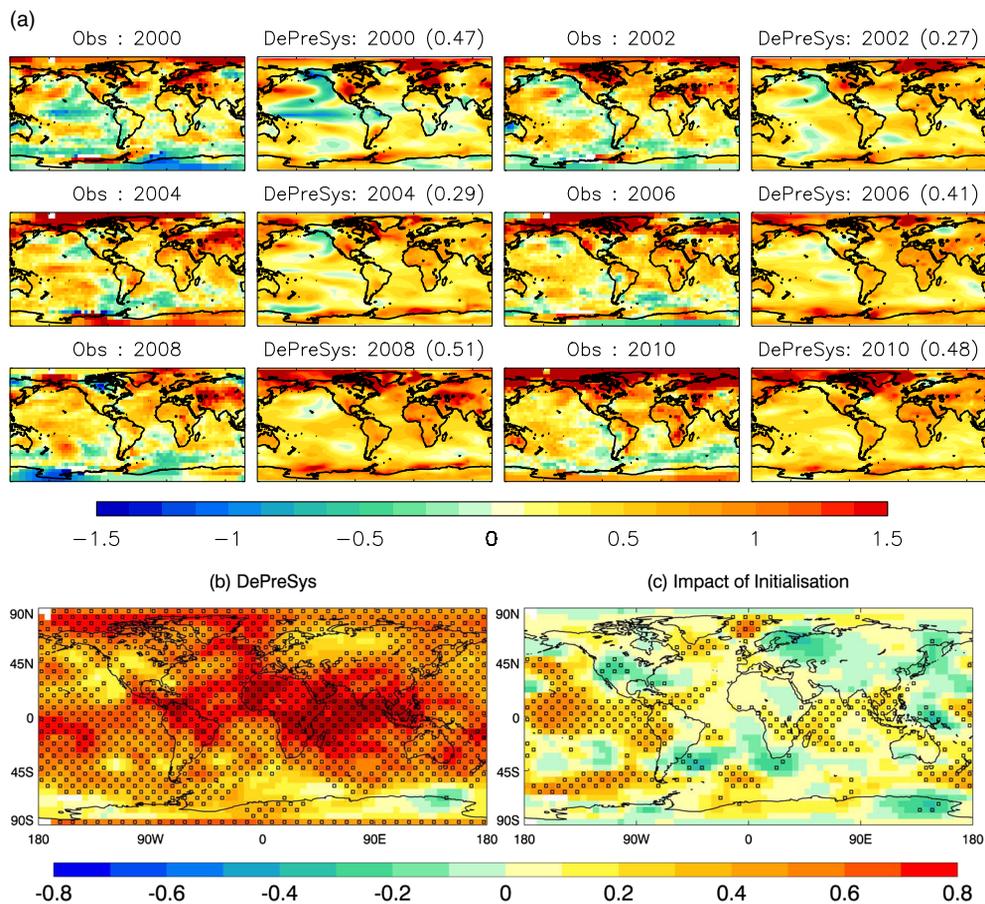
Figure 4. (a) Evaluation of dynamical hindcasts. Example patterns over 2000–2011 of observed surface temperature anomalies from a 1961 to 1990 average (left panels for each year) and corresponding hindcast surface temperature anomalies (right panels for each year). Numbers in brackets are spatial correlations of the hindcast with observations. Observations are averages of the HadCRUT3, GISS, and NCDC data sets; values are plotted where there is at least one value from any data set. (b) Temporal correlation between hindcast and observed temperature from 50 hindcasts over 1960–2009. (c) Impact of initialization on temporal correlations calculated as the total correlation from b minus that of parallel un-initialized hindcasts, driven by changes in anthropogenic and natural radiative forcings alone. Stippling in b and c shows correlations significant at the 5% level based on 1000 bootstrap resamplings.

References

Brohan, P., J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850, *J. Geophys. Res. (Atmos.)*, *111*, D12106, doi:10.1029/2005JD006548.

Folland, C. K., D. E. Parker, A. W. Colman, and R. Washington (1999), Large scale modes of ocean surface temperature since the late nineteenth century, in Beyond El Niño: Decadal and Interdecadal Climate Variability, Chapter.4, edited by A. Navarra, p. 374, Springer, New York.

Fröhlich, C., and J. Lean (1998), The sun's total irradiance: Cycles, trends and related climate change uncertainties since 1976, *Geophys. Res. Lett.*, *25*, 4377–4380.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Climate Dyn.*, *16*, 147–168.

Gray, L. J., et al. (2010), Solar influences on climate, *Rev. Geophys.*, *48*, RG4001, doi:10.1029/2009RG000282.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis (2010), Probing the fast and slow components of global warming by returning abruptly to pre-industrial forcing, *J. Clim.*, *23*, 2418–2427, doi:10.1175/2009JCLI3466.1.

Johns, T. C., et al. (2003), Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios, *Clim. Dyn.*, *20*, 583–612, doi:10.1007/s00382-002-0296-y.

Knight, J., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann (2005), Natural variations in the thermohaline circulation and future surface temperature, *Geophys. Res. Lett.*, *32*, L20708, doi:10.1029/2005GL024233.

Lawson, N. (2008), An appeal to reason: A cool look at global warming, Duckworth Overlook, London.

Meinhausen, M., et al. (2011), The RCP greenhouse gas concentrations and their extension from 1765 to 2300, *Clim. Chang.*, *109*, 213–241, doi:10.1007/s10584-011-0156-z.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, *J. Geophys. Res. (Atmos.)*, *117*, D08101, doi:10.1029/2011JD017187.

Parker D. E., H. Wilson, P. D. Jones, J. R. Christy, and C. K. Folland (1996), The impact of Mount Pinatubo on worldwide temperatures, *Int. J. Climatol.*, *16*, 487–497.

Parker, D. E., et al. (2007), Decadal to interdecadal climate variability and predictability and the background of climate change, *J. Geophys. Res. (Atmos.)*, *112*, D18115 doi:10.1029/2007JD008411.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack (1993), Stratospheric aerosol optical depths, 1850–1990, *J. Geophys. Res. (Atmos.)*, *98*, 22,987–22,994.

Simmons, A., K. Willett, P. Jones, P. Thorne, and D. Dee (2010), Low frequency variation in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalysis and monthly gridded observational data sets, *J. Geophys. Res.(Atmos.)*, *115*, D01110, doi:10.1029/2009JD012442.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model. *Science*, *317*, 796–799, doi:10.1126/science.1139540.

Smith, D. M., and J. M. Murphy (2007), An objective ocean temperature and salinity analysis using covariances from a global climate model, *J. Geophys. Res., (Oceans)*, *112*, C02022, doi:10.1029/2005JC003172.

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

Thompson, D. W., J. M. Wallace, J. J. Kennedy, and P. D. Jones (2010), An abrupt drop in Northern Hemisphere sea surface temperature around 1970, *Nature*, *467*, 444–447, doi:10.1038/nature09394.

Vernier, J.-P., et al. (2011), Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, *Geophys. Res. Lett.*, *38*, L12807, doi:10.1029/2011GL047563.