Simulation models are widely employed to make probability forecasts on seasonal to annual timescale and increasingly on longer time-scales. While dynamical models are often expected, in principle, to represent purely empirical models, that claim must be established empirically for any given generation of model; direct comparison of the forecast skill of any dynamical and empirical model must provide information on its performance that is not available in model intercomparisons. More importantly, the blending of forecasts from both sources can lead to better operational forecasts. Decadal scale forecasts can also reveal the space and time scales on which model simulation models exploit their physical basis effectively, perhaps indicating the origin of their weaknesses. The skill of decadal and seasonal probabilistic hindcasts for global and regional mean temperatures from the ENSEMBLES projects is interpreted in this context. Physically inspired empirical models are shown to display probabilistic skill comparable to that of today’s state-of-the-art simulation models as well as that of the multimodel ensemble. The inclusion of empirical models (blending) with simulation models produces significant improvements forecasts. As the size of building or running empirical model is negligible compared to large simulation models, it is suggested that the direct comparison of simulation models with empirical models is not feasible. While simulation models based on physical principles can be expected to reproduce empirical models whenever the timescale allows, and that blending simulation models with empirical models become a regular component of seasonal and decadal scale forecasting systems.

Seasonal Forecasts

The ENSEMBLES multimodal-ensemble experiment for seasonal forecasting compares the global coupled atmosphere-ocean climate models (Further details of the ENSEMBLES experiments can be found in [5]). Hindcast simulations considered here were launched on the first day of February, March, August and November each year over the 46-year period from 1960 to 2005 for the Nino3.4 region (each model consists of a nine-member initial condition ensemble). Probability forecasts are generated from the ENSEMBLES simulations via kernel dressing and are blended with climatology to produce probabilistic forecasts (for a full description see Appendix I). Figure 1 shows an example of the kernel-dressed and blended probabilistic forecast distributions for a value (over the period 1995-2006) of the IPSL/CERFACS hindcast simulations from the ENSEMBLES project launched in November.

Decadal Forecasts

A set of decadal simulations from the ENSEMBLES-Based Pre-Edition of Climate Change and Their Impacts (ENSEMBLES) experiment ( Bengtsson and S. Solomon, 2006; Dufresne et al., 2006) to perform phase 5 of the Coupled Model Intercomparison Project (CMIP5) decadal simulations (Taylor et al., 2009). The Ensembles-MODEL, each ensemble simulation consisted of only three members launched in 1960, 1970 and 1980. Figure 4. Illustrates the 2-year running mean of simulated global-mean temperature from the four simulation ensembles of the ENSEMBLES project over the full set of hind cast data. Observations from the Hadley Centre/Climatic Research Unit, version 3 (HadCRUT3) and the 40 year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis 40 year (ERA40) are shown for comparison. HadCRUT3 is used as the verification dataset. The evaluation of the multimodel ensemble was carried out at the annual mean level. The ensemble mean output is seen to be the average of the ensemble means at both times of absolute values, as well as in dynamic.

Appendix I: From Simulation to a PDF

Probabilistic seasonal forecasts based on the ENSEMBLES simulation models clearly outperform climatologically probabilistic forecasts in many cases. The fact that empirical-based probabilistic forecasts are provided for some of the skill of operational systems can be enhanced with information from the richer empirical models. The quality of decadal probability forecasts from the ENSEMBLES simulation models has been compared with that of reference forecasts from several empirical models. In general, the seasonal ENSEMBLES simulation models are superior to the empirical models across the range of lead-time frames. To 10 years. The ENSEMBLES probabilistic forecasts often provide up to 4 life more information (or 2 times more probability mass) on the observed outcome than the empirical forecasts. Probability forecasts from empirical models are often more accurate in climate simulations. The combined skill of the ENSEMBLES probabilistic hindcasts and empirical models suggest that the empirical models will be used for blending with simulation model ensembles. It also allows interpretation of the combined ensemble to which operational forecasts can be closely related. Blending empirical model simulations with the climate simulations based on the simulation of the Earth system and can thereby be expected to provide robust, reliable machines (and, of course, operate empirically) on longer time scales. The component of near-term climate predictions from Earth simulation models with those from dynamic climate empirical models provides a useful benchmark as the simulation models improve in the future.

Appendix II: Evaluate probabilistic forecasts

The performance of forecast distributions is evaluated primarily using the key score Ignorance Score [2]. The Ignorance Score is defined by:

$$I(p, x; t) = \sum_{i=1}^{K} u_i p_i - \sum_{i=1}^{K} u_i p_i q_i$$

where $p_i$ is the density function generated by the ensemble and $q_i$ is the estimate of climatological density. To perform a near-term climate simulation model, a combination of the kernel density estimate (KDE) kernel width $\alpha$ is assigned to the model. We do these three parametres simultaneously by optimizing the Ignorance score, introduced below, by least one cross validation (see [5] for a full description).