International Journal of Forecasting 34 (2018) 336–338

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Discussion

Commentary on "Data-based mechanistic modelling and forecasting globally averaged surface temperature"

P. Geoffrey Allen

Department of Resource Economics, University of Massachusetts, USA

Peter Young's main aim is to provide a statistical model of how changes in total radiative forcing cause changes in average annual global surface temperature (GST) and use this for short- to medium-term forecasting of GST. This is clearly a highly aggregated variable, but is the measure of climate that has attracted the most political attention.

GST is usually reported as an anomaly, that is relative to a base. For HadCrut4, the data series used in this study, the base is the average global temperature from 1961–90. Outputs from General Circulation Models (GCMs), the main source to date of GST projections, are reported in degrees Kelvin and in order to compare them with Young's forecasts these must be converted to anomalies using the 1961–90 climatology. Young's choice of total radiative forcing (TRF) as exogenous variable is less common than using carbon dioxide concentration. TRF, despite its name is a net measure of the incoming solar radiation less that radiated back into space. A larger TRF means that the earth absorbs more radiation leading to a warmer planet. Greenhouse gases, principally carbon dioxide, reduce the energy radiated back, while aerosols (e.g., volcanic ash, dust, industrial pollutants) work in the opposite direction. Therefore, TRF is a convenient variable to capture the effect of all components on the earth's energy balance. In practice, since greenhouse gas concentrations are highly negatively correlated with aerosol concentrations and the reduction in re-radiation caused by lesser greenhouse gases such as methane and oxides of nitrogen has closely matched the increase caused by aerosols, there is probably little difference between using TRF or CO2 concentration as the exogenous variable for forecasting, though the historical variations in TRF because of volcanic eruptions do affect parameter estimates and through them the values forecast.

Young uses a continuous-time transfer function as the core of his model. I have discovered only one other instance of this model, and that by someone at the same university (Li & Jarvis, 2009). Transfer functions of the discrete-time variety have had a brief and limited appeal in business forecasting. The distinction is that instead of lagged variables, continuous-time transfer functions consist of first-order and higher derivatives. These variables must be created from the observed data series by means not specified in the paper but included in the estimation algorithms (see Young, 2015). Although the deterministic component of the transfer function is continuous, the noise component is discrete, producing a hybrid function.

Young first compares the output of a GCM (the Princeton University Geophysical Fluid Dynamics Lab GFDL15a) with the output from a MAGICC model. MAGICC is an emulator that incorporates a detailed list of forcing variables (greenhouse gases and aerosols) and models temperature responses in a simplified model of the earth (which is treated as two boxes, northern and southern hemisphere). MAGICC has a number of parameters that can be adjusted (though it is large enough that they are not easily optimized in a statistical sense); the MAGICC model used in the comparison was not adjusted. That the simulations need 4,000 years to approach equilibrium shows that the earth's climate dynamics are believed to have extremely long-lasting components. Young then compares the GFDL output with the deterministic part of a first-order continuous transfer function and finds that the simple transfer function fits about as well as the MAGICC emulator. Interestingly, this suggests that the earth's response to increased CO2 concentration is governed by one or two dominant modes, in which case, simple statistical models should be sufficient to forecast aggregate climate changes.

The main part of the paper is an unobserved components model of trend, cycle, response to forcing and noise. Trend is a temperature shift in addition to that resulting from a trending forcing variable and is here set to zero as this is not required. The response component is a first-order hybrid transfer function of two input variables, total radiative forcing and a constant to deal with an initial

E-mail address: allen@resecon.umass.edu.

https://doi.org/10.1016/j.ijforecast.2017.10.001

non-zero condition. The residual from the transfer function is treated as a compound wave form. Spectral analysis suggests there are cycles with 13 peaks which constitute the 'quasi-cyclical' component whose frequency and phase characteristics can change a little over time. These two components are optimized together. The complex nature of this quasi-cycle is illustrated in the bottom panel of Young's Figure 4.

In order to forecast GST, forecast values of the forcing variable are needed. Young uses a Dynamic Harmonic Regression model based on an AR(7) spectrum, with a trend, to forecast values of TRF. It could be regarded as the forecast for a no-policy-change situation. The large dips in the historical series show the substantial, but short-term, impact of major volcanic eruptions. This contrasts with the usual approach to both decadal and longer-term forecasting of GST which is to use a Representative Concentration Pathway (RCP). That includes concentrations of greenhouse gases and aerosols, except volcanic ash (on the grounds these are unforecastable). The RCP projections are assumed to represent the effects of different energy policies and in that sense could be considered conditional forecasts. For decadal forecasts, it makes little difference which RCP is used, since they are similar in the first few years but begin to diverge after about 2050. In Young's case, the quasi-cycle is projected forward using the previously estimated parameters. It can be smoothed by using fewer than the 13 components, which seems advisable in this forecasting context. Forecasting of the response component is done by sampling the output of the continuous-time transfer function and updating the forecast by Kalman filtering. Adding the two forecast components gives the final forecast.

Young then makes two forecast comparisons and proposes one model extension. First, he compares the forecasts from his data-based model with those from a discrete-time transfer function using a Box-Jenkins approach. Based on either MAPE or MSE, the 10-year rolling forecasts from 1980 to 2000 from the DBM are always more accurate than the forecasts from a discrete-time transfer function which are almost always more accurate than persistence or naïve no change. It would be interesting to compare just the continuous-time and discrete-time transfer functions. My suspicion is that it is the addition of the quasi-cycle component that makes the difference, so including it in one case but not the other is hardly a fair comparison.

Second, Young compares out-of-sample forecasts from 2002-2010 with actual temperature anomalies and with bias-corrected forecasts from three climate models that contributed to the Fifth Coupled Model Intercomparison Project (CMIP5). In making decadal forecasts climate modelers have become aware of issues not apparent in century-scale projections: the impact of initial conditions and the distinction between a model's climatology and actual observations. These manifest themselves as model bias, which needs to be corrected. One impediment to good bias correction is the limited number of errors from decadal forecasts from which to make the correction. As shown in Young's Figure 10, the individual GCM forecasts and their average (the solid line) overforecast the actual temperature observations while the DBM forecasts are much closer.

Third, Young proposes the addition of other explanatory variables, suggesting the Atlantic Multidecadal Oscillation as a possibility. This echoes the approach of other statistical and regression models which have used the El Nino-Southern Oscillation (ENSO) and the Pacific Decadal Oscillation. The earth's dynamics are partly understood to the extent that there are fast processes that are over in a matter of minutes to hours and slower processes, some of which take thousands of years to play out, most notably heat transfers in the deep ocean. Using these oscillation variables as summarizing complex processes is analogous to the use of jointly dependent variables in simultaneous equations estimation. And, of course, for forecasting, they also need to be forecast.

Finally, Young compares equilibrium climate sensitivity (ECS) from his DBM approach when emulating the GFDL model and directly from observations. In the first case he estimates an ECS of 4.3°C and in the second case of 2.6°C with a 95% confidence interval of 2.1–3.0°C, showing the flexibility of the approach. ECS is the global average surface temperature increase caused by an immediate doubling of the CO2 concentration above the pre-industrial level. In GCMs this can be obtained directly from the experiment of a step-increase in CO2. In statistical analyses, as in Young's paper, it is calculated from estimated parameters, while in MAGICC it is a directly-controlled parameter. In emulating the GFDL GCM, Young's estimate is on the high side of reported values from GCMs. (ECS for five GFDL models in the CMIP5 and the earlier CMIP3 range from 2.4 to 4.0°C .) For the 42 GCMs in the two Coupled Model Intercomparison Projects reporting a value for ECS, the mean is just over 3.2°C, while the range of values, from 2.1 to 4.6°C indicates the uncertainty of this measure. These are central estimates from an ensemble of simulations by each model and do not therefore measure a prediction interval. In sharp contrast, a non-exhaustive list of 12 recent statistical studies using energy balance, regression or simple models has a mean ECS of 1.8°C, a median of 2.0°C, and a range of 0.7–2.6°C. Most have a confidence interval, where reported, noticeably wider than Young's, suggesting that his approach can deliver a more precise forecast than other methods.

Of course, an instantaneous doubling of CO2 concentration is an artificial construct. But ECS is a useful summary measure of the response of temperature to CO2 concentration change according to a particular climate model or statistical analysis. The striking finding to date is that almost all the analyses that rely on an actual temperature series, including reconstructions that go back thousands of years, find an ECS lower than all but one or two GCMs. ECS is a very uncertain measure but the weight of evidence shows a clear distinction between the two approaches. It is hard to argue against the findings of the statistical analyses. Young's analysis shows that his method is capable of matching GCM values of ECS on GCM output and of other statistical analyses when using actual observations.

Complex General Circulation Models are undoubtedly the way forward for wide-ranging studies of climate change since they can capture so many of the effects that have societal implications: regional temperature extremes, precipitation changes, ice melting, and so on. Yet their very complexity makes them suspect. Parameter adjustment or "tuning" is a compromise and when detailed outputs are aggregated, problems are revealed, for example, when decadal forecasts of GST (of which there are a large number from the CMIP5 experiments) are compared with actual observations. Simpler models provide a benchmark, particularly for decadal forecasting, against which the GCMs can be measured. The key question, with current knowledge, is which of the two divergent predictions should be believed. Is it the range of values from GCMs, complex models based on the laws of fluid mechanics which until the era of decadal forecasts had never been tested against actual observations? Or is it the statistical models of various kinds that are based on actual observations, though with limited data, and whose specifications are suspect and in some cases surely inadequate? For example, some statistical models use CO2 emissions as explanatory variable while it is CO2 concentration that directly causes surface temperature rise, albeit with some lag.

Does the continuous-time transfer function have a future? Its parameters are not a function of the sampling interval, though for forecasting at the same time interval this is not really an issue. It appears to provide an improvement over the discrete time transfer function. More important, we need to consider the addition of a quasi-cycle to the model when there are time-varying oscillations, since its addition appears to be the major source of improvement in forecast accuracy. There are probably such time-varying oscillations in economic systems. The recent Great Recession revealed that the business cycle is still alive and well and was not forecast by mainstream economists. We need more forecasts from this approach and comparisons with discrete-time methods to see if the promise revealed in Peter Young's paper can be borne out. He makes a compelling case for further investigation.

References

Li, Sile, & Jarvis, Andrew (2009). Long run surface temperature dynamics of an A-OGCM: the HadCM3 $4 \times$ CO2 forcing experiment revisited. *Climate Dynamics*, *33*(6), 817–825.

Young, P. C. (2015). Refined instrumental variable estimation: Maximum likelihood optimization of a unified Box-Jenkins model. *Automatica*, *52*, 35–46.

P. Geoffrey Allen is Emeritus Professor in the Department of Resource Economics. He has worked on various aspects of econometric forecasting for most of his career, publishing a definitive appraisal of forecasting in agriculture in the IJF. His current research interest is in decadal climate forecasting.