A distance correlation ratio of predictable components

Rick Danielson (rickedanielson@gmail.com)
Fisheries and Oceans Canada, Halifax, Nova Scotia

Introduction

The ratio of predictable components (RPC) is a useful metric of forecast skill (Kumar et al. 2014; Eade et al. 2014), although it begs the question of whether one should expect an ensemble mean to covary more with observations than with its own ensemble members. An implicit assumption is that the forecast model is independent of observations, but even if the forecast model is not assimilative, there remains the possibility that errors are correlated, or in other words, that the observation and forecast model *measures* [in the context of a simple *measurement model* (Siegert et al. 2016)] are still not independent. A new correlation method has been developed by Székely et al. (2007) and Székely and Rizzo (2009) that allows this statistical independence assumption to be tested. We compare the two RPC values for a single raw ensemble simulation using the Lorenz 1963 modelling framework of Mayer et al. (2021).

Results

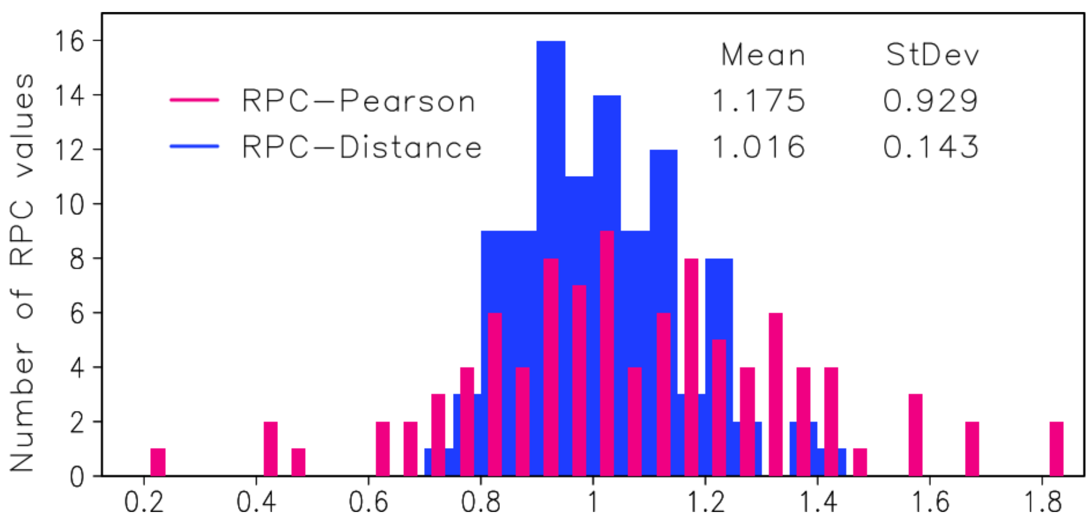

Figure 1: Distributions of the ratio of predictable components (RPC) for 100 40-member ensemble hindcasts of the Lorenz model (Mayer et al. 2021), where RPC is computed as a ratio of (red) Pearson correlations (Scaife and Smith 2018) and (blue) distance correlations (Székely et al. 2007; Székely and Rizzo 2009). The bin interval is 0.05 and the mean and standard deviation of the two distributions are included.

It is convenient to calculate a distance correlation RPC without the added concern of timeseries autocorrelation. Thus, instead of the monthly and seasonally averaged timeseries that is the focus of Mayer et al. (2021), we start with one of their raw ensemble simulations. For each of 100 different initial conditions on the Lorenz attractor, a noisy observational simulation and a 40-member ensemble is used to calculate RPC using either a ratio of Pearson or distance correlations (the RPC numerator is the correlation of observations with ensemble-mean, and the denominator is the average correlation over all ensemble members with the mean, excluding that member). In order to avoid autocorrelation, 20 samples at 100 time intervals are taken from each 2001-unit perturbed raw timeseries.

Figure 1 shows one of the underconfident (i.e., large initial spread) ensemble experiments of Mayer et al. (2021), which yields a mean RPC (Pearson) of greater than one and a large RPC standard deviation as well. Interestingly, the RPC (distance correlation) has a mean value closer to one, and hence, is more consistent with expectations (Kumar et al. 2014; Eade et al. 2014). The RPC variance is also much reduced.

Summary

It may be difficult to define predictability, and thus any *linear* measure of it, but perhaps a distance correlation RPC is also a viable diagnostic of forecast model skill. The physical basis for assuming that measures are independent is sound, although measurement models themselves are typically simple, and may implicitly impose a dependence. Strictly speaking, we do not consider a forecast ensemble that is marginally calibrated (Siegert et al. 2016), but it seems notable that under this assumption, RPC is approximately β^{-1} (i.e., inverse of the multiplicative parameter of linear agreement between model and observations). However, it is difficult to anticipate a value of β between the limits of ordinary and reverse linear regression. For a given value of distance correlation between 0 and 1, it seems that any value of β is possible (Edelmann et al. 2021).

Acknowledgements

M. Rizzo and G. Szekely are thanked for providing the distance correlation R package (energy). B. Mayer is thanked for providing the Lorenz model framework (https://doi.org/10.5281/zenodo.4418958), and J. Mecking is thanked for drawing attention to the RPC.

References

Eade, R., D. Smith, A. A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, doi:10.1002/2014GL061146.

Edelmann, D., T. F. Móri, and G. J. Székely, 2021: On relationships between the Pearson and the distance correlation coefficients. *Stat. Prob. Lett.*, **169**, 1–6, doi:10.1016/j.spl.2020.108960.

Kumar, A., P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, doi:10.1175/MWR–D–13–00287.1.

Mayer, B., A. Düsterhus, and J. Baehr, 2021: When does the Lorenz 1963 Model exhibit the signal-to-noise paradox? *Geophys. Res. Lett.*, **48**, 1–10, doi:10.1029/2020GL089283.

Scaife, A. A., and D. Smith, 2018: A signal-to-noise paradox in climate science. *npj Clim. Atmos. Sci.*, **1**, 1–8, doi:10.1038/s41612–018–0038–4.

Siegert, S., D. B. Stephenson, P. G. Sansom, A. A. Scaife, R. Eade, and A. Arribas, 2016: A Bayesian framework for verification and recalibration of ensemble forecasts: How uncertain is NAO predictability? **29**, 995–1012, doi:10.1175/JCLI–D–15–0196.1.

Székely, G. J., and M. L. Rizzo, 2009: Brownian distance covariance. *Annals of Applied Statistics*, **3**, 1236–1265, doi:10.1214/09AOAS312.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov, 2007: Measuring and testing dependence by correlation of distances. *Annals of Statistics*, **35**, 2769–2794, doi:10.1214/009053607000000505.