



## Reply to Comments on “Multiyear Predictions of North Atlantic Hurricane Frequency: Promise and Limitations”

GABRIEL A. VECCHI,\* RYM MSADEK,<sup>+</sup> WHIT ANDERSON,\* YOU-SOON CHANG,<sup>#</sup> THOMAS DELWORTH,\* KEITH DIXON,\* RICH GUDGEL,\* ANTHONY ROSATI,\* BILL STERN,\* GABRIELE VILLARINI,<sup>@</sup> ANDREW WITTENBERG,\* XIASONG YANG,<sup>+</sup> FANRONG ZENG,\* RONG ZHANG,\* AND SHAOQING ZHANG\*

\* NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

<sup>+</sup> NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, and University Corporation for Atmospheric Research, Boulder, Colorado

<sup>#</sup> Department of Earth Science, Kongju National University, Kongju, South Korea

<sup>@</sup> IHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa

(Manuscript received and in final form 2 July 2013)

In our original paper (Vecchi et al. 2013, hereafter V13), we stated “the skill in the initialized forecasts comes in large part from the persistence of the mid-1990s shift by the initialized forecasts, rather than from predicting its evolution.” Smith et al. (2013, hereafter S13) challenge that assertion, contending that the Met Office Decadal Prediction System (DePreSys) was able to make a successful retrospective forecast of that shift. We stand by our original assertion and present additional analyses using output from DePreSys retrospective forecasts to support our assessment.

S13 argue that the physical basis for multiyear hurricane prediction comes from (i) multiyear predictability of North Atlantic subpolar gyre temperature (SGT) and (ii) the impact of SGT changes on hurricane activity, thereby creating skillful multiyear predictions for hurricanes. We agree that skillful multiyear predictions for North Atlantic SGTs have been made with DePreSys and other systems (e.g., Robson et al. 2012; Yeager et al. 2012), including the GFDL prediction system (Yang et al. 2013; R. Msadek et al. 2013, unpublished manuscript). We also agree that there are physically plausible hypotheses by which SGT changes could influence hurricane activity (e.g., Sutton and Hodson 2005; Zhang and Delworth 2006; Kang et al. 2008; Dunstone et al. 2011), although the magnitude of this influence is not clear.

However, we disagree that this multiyear predictive skill for SGTs has in fact produced skillful multiyear

retrospective forecasts of hurricane activity, apart from the effect of the persistence of the shift. Here we provide additional support for our assessment through the analysis of retrospective forecasts of hurricane activity from DePreSys (Smith et al. 2010), in addition to the work originally presented in V13. Output from the DePreSys retrospective forecasts was kindly provided by Doug Smith. The data used in the present reply and in S13 differ from those in the original V13, with V13 using an SST-based index as a proxy for hurricane activity and S13 and the present study counting low pressure systems in a dynamical model. We have conducted analyses using the DePreSys hurricane retrospective forecasts for lead years 1–5, 2–6, and 3–7; all leads yield similar results. In Fig. 1a we plot the number of forecast hurricanes for lead years 1–5 (vertical axis) versus the number of observed hurricanes (horizontal axis; 5-yr running mean centered on retrospective forecast year 3), with one symbol for each initialization year. A set of perfect retrospective forecasts would form a straight line with a 1:1 slope (i.e., for each year the observed and retrospective forecast hurricane counts would be identical). Retrospective forecasts initialized in the “preshift period” (blue circles) show an essentially random scatter, with no significant correlation between observed and forecast counts (the correlation is 0.07). This indicates no skill. For the postshift period (red circles), the correlation is  $-0.47$ , again indicating no skill. Thus, for the periods before and after the shift, there is no demonstrated skill in multiyear retrospective forecasts of hurricane counts, as was argued in V13 using a different hurricane prediction methodology.

Corresponding author address: Gabriel A. Vecchi, NOAA/GFDL Princeton, NJ 08540.  
E-mail: gabriel.a.vecchi@noaa.gov

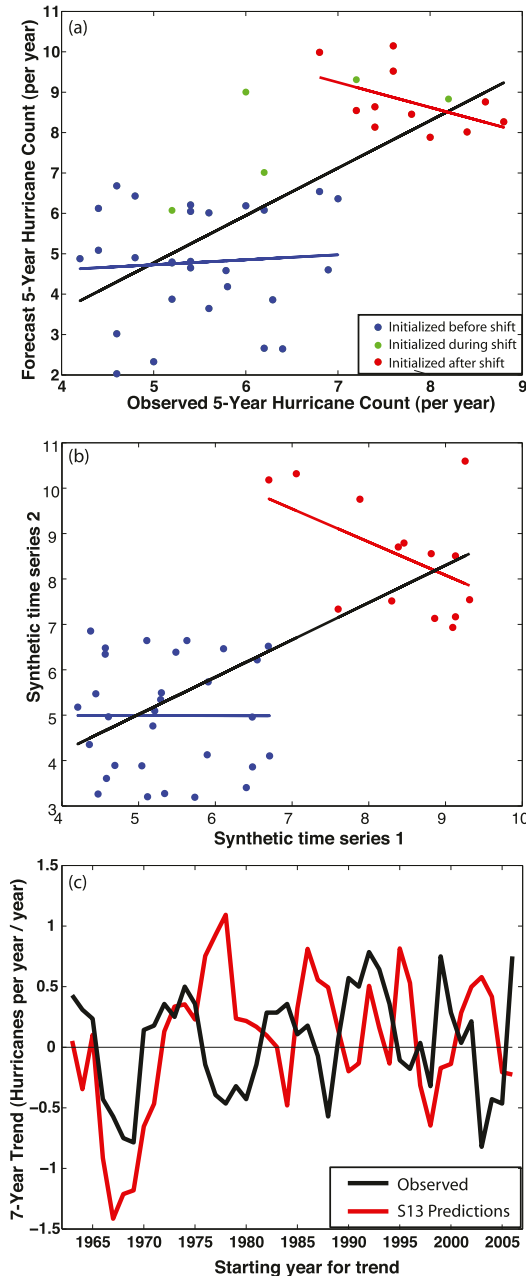


FIG. 1. (a) Scatterplot showing observed and forecast hurricane counts. Each circle corresponds to a retrospective forecast of hurricane counts starting from a particular initialization time. For each circle, the value along the y axis corresponds to the forecast count averaged over lead years 1–5 and the value along the x axis corresponds to the 5-yr running mean of the observed hurricane count centered at the midpoint of the forecast (lead year 3). A suite of perfect retrospective forecasts would correspond to a set of circles with a 1:1 slope. The blue circles are for retrospective forecasts initialized in years 1962–89 (before the mid-1990s shift), the green circles are for retrospective forecasts initialized in years 1990–94 (forecast crosses the mid-1990s shift), and the red circles are for retrospective forecasts initialized in 1995 and later (after the mid-1990s shift). The black line corresponds to a linear regression fit using all initialization years, while the blue (red) line corresponds

When all years are considered (regression shown by black line in Fig. 1a), the correlation between retrospective forecast and observed counts is 0.67; this appears encouraging. However, it is clear from the figure that this relatively large correlation coefficient results entirely from the mid-1990s shift in Atlantic hurricane frequency; there are two distinct populations of points, each without any forecast skill, so that a regression/correlation using all points primarily reflects the difference in the mean between the groups. For a perspective on this result, we examine pairs of synthetic time series constructed as sequences of random numbers, scaled to have the same mean and standard deviation as the observational time series, such that the underlying correlation between such pairs of time series is zero. If we now impose a shift in each time series similar in magnitude and timing to the observed mid-1990s shift, the average correlation between such pairs of time series (averaged over many realizations) increases from approximately zero to 0.70. The scatterplot for a representative pair of such time series is shown in Fig. 1b. The relatively large correlation is entirely due to the presence of the shift; there is no correlation without the shift. We suggest that a similar process operates with regard to correlations between retrospective forecast and observed hurricane counts and that the computed correlation over the entire record reflects the shift.

Since any indication of potential forecast skill comes from the mid-1990s shift, it is crucial to determine whether there was skill in retrospectively predicting the shift itself. To make an assessment of prediction skill of the shift, we must focus on a metric that evaluates the evolution of retrospective forecasts over the various leads. We suggest that comparing the 7-yr trends in hurricane counts derived from the retrospective forecasts over lead years 1–7 with the corresponding 7-yr trends from observations provides such a metric (Fig. 1c). In observations, there were multiyear periods where the 7-yr trends tended to be persistently positive or negative, with the largest being an upward shift in the mid-1990s; there are also multiyear

←

to a regression line using only initialization years prior to (after) the mid-1990s shift. (b) As in (a), but using the output from two synthetic time series derived from random numbers but scaled to have the same mean and variance as the observed time series, along with a similar shift in the time series. Blue circles are times before the shift, and red circles are times after the shift. We show one representative realization out of 1000 realizations conducted. (c) Shown are 7-yr trends in hurricane counts computed over lead years 1–7 from the retrospective forecasts (red) and corresponding observations (black). The trend (slope of the least squares linear fit) is plotted at the year corresponding to lead year 1 (e.g., trends for 1963–69 are plotted at year 1963).

periods where the retrospective forecasts tend to increase/decrease consistently for various initialization years. However, there is little correspondence between the observed history of 7-yr trends and those across lead years 1–7 of the retrospective predictions, with the correlation coefficient being 0.11. We have performed similar trend calculations using all available trend lengths from the retrospective forecasts and, while the details differ between various trend lengths, we find no significant skill for any trend length.

Focusing on the time frame of the observed mid-1990s shift, there are modest positive trends in the retrospective forecasts. However, they do not stand out from a number of other such retrospective forecast trends over the entire record. In contrast, the observed trends in the early 1990s are the largest in the record. There are a number of other periods (1970s, 1980s, and early 2000s) where the retrospective forecasts have large trend values that are not seen in the observed trends, with the most recent occurring at a time when observations showed a substantial decrease. Much of the nominal positive correlation between the observed and retrospectively forecast trends comes from the early part of the record, during which the retrospective forecast system did indicate a rapid decrease in hurricane frequency initialized in the middle to late 1960s, as was seen in the observations (the correlation over the first 12 yr is 0.91 and  $-0.4$  thereafter). In this reply, we have focused on the analysis of the hurricane predictions described in S13, which were not available to us at the time of the original V13 manuscript was prepared. We note that the analyses described in this reply produce effectively the same results when applied to the hurricane predictions used in V13.

A crucial point is that, even with 50 yr of data, we are dealing with a very small sample size (as we highlighted in V13), in which a single event (the mid-1990s shift) is crucial to any statements about skill. With such a small sample it is very difficult to evaluate whether there is in fact robust skill. Yet, the claim in S13 that the system skillfully forecast the mid-1990s shift is not supported by this analysis. While there are small positive values of the forecast trend in the early 1990s, these do not stand out from the record and must be interpreted in light of other periods with even larger positive trends in the forecasts that did not occur in observations.

S13 speculates that the presence of other factors, such as unpredictable volcanic eruptions, could account for some of the differences between the retrospective forecast and observed trends. Without a rigorous evaluation of these possibilities through additional retrospective forecasts that examined the impact of such processes on skill, there is no

way to assess the validity of such suggestions. We can only assess the retrospective forecasts that were made.

These results do not imply that skillful multiyear predictions of Atlantic hurricane activity are not possible. Rather, they show that skill has not (yet) been demonstrated, and they do not preclude the possibility of skill using systems not considered here. There is clearly skill for multiyear retrospective forecasts of North Atlantic SGT. Further, there are postulated physical mechanisms linking subpolar gyre conditions to hurricanes. Thus, this potential source of predictive skill for hurricane activity clearly merits further exploration. However, our assessment is that the prediction systems evaluated to date do not yield skillful multiyear retrospective forecasts of hurricane activity, with the exception of skill that arises from the existence and persistence of the mid-1990s shift. A fundamental challenge to assessing skill is the limited observational record, providing very few cases against which to evaluate any predictive system.

#### REFERENCES

- Dunstone, N. J., D. M. Smith, and R. Eade, 2011: Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophys. Res. Lett.*, **38**, L14701, doi:10.1029/2011GL047949.
- Kang, S. M., I. M. Held, D. M. W. Frierson, and M. Zhao, 2008: The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *J. Climate*, **21**, 3521–3532.
- Robson, J. I., R. T. Sutton, and D. M. Smith, 2012: Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophys. Res. Lett.*, **39**, L19713, doi:10.1029/2012GL053370.
- Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife, 2010: Skillful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.*, **3**, 846–849, doi:10.1038/ngeo1004.
- , and Coauthors, 2014: Comments on “Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations.” *J. Climate*, **27**, 487–489.
- Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309**, 115–118, doi:10.1126/science.1109496.
- Vecchi, G. A., and Coauthors, 2013: Multiyear predictions of North Atlantic hurricane frequency: Promise and limitations. *J. Climate*, **26**, 5337–5357.
- Yang, X., and Coauthors, 2013: A predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system. *J. Climate*, **26**, 650–661.
- Yeager, S. G., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: Late twentieth-century North Atlantic Ocean heat content. *J. Climate*, **25**, 5173–5189.
- Zhang, R., and T. L. Delworth, 2006: Impact of Atlantic multi-decadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.*, **33**, L17712, doi:10.1029/2006GL026267.