Forecast for the 2013 All-Indian Summer Monsoon Rainfall

Timothy DelSole[1,2], Michael Tippett[3,4], Jyothi Nattala[2]

Issued: June 12, 2013

Summary

We predict that the all-India June-September monsoon rainfall for 2013 is more likely to be below normal than above normal, relative to the 1871-2010 climatology. A distinctive feature of this prediction is that it is based on a regression model that uses ENSO indices *predicted* by the NCEP Climate Forecast System (version 2; CFSv2). Specifically, for 2013, the CFSv2 model initialized on May 1 and May 6 predicts that June-September sea surface temperatures in the central equatorial Pacific will be about 0.27°K above the 1999-2010 model climatology. Based on this predicted sea surface temperature, a regression prediction of all-Indian monsoon rainfall gives an approximately Gaussian distribution with a mean of 95% of the 1871-2010 long term mean, and a standard deviation of 10% of the 1871-2010 long term mean. For reference, the mean of 1999-2010 was 96% of the long term mean, so the mean of the probabilistic prediction is not atypical of the most recent period. This is an experimental forecast intended for information purposes only; it is not intended to replace official advisory, forecast, or warnings issued by government agencies.

[1]George Mason University, 4400 University Dr Fairfax, VA 22030
[2]Center for Ocean-Land-Atmosphere Studies, Calverton, MD 20705
[3]International Research Institute for Climate and Society, Palisades, NY
[4]King Abdulaziz University, Jeddah, Saudi Arabia

1 Introduction

This document summarizes a forecast of the 2013 All-Indian Summer Monsoon Rainfall based on information that is available in early May. The basis of this forecast is the recent discovery that coupled atmosphere-ocean models can predict monsoon rainfall with statistically significant skill (Rajeevan et al., 2011; DelSole and Shukla, 2012). Evidence of this fact is shown in fig. 1, taken from DelSole and Shukla (2012), which shows the correlation skill between predicted and observed monsoon rainfall from the ENSEMBLES seasonal forecasts. In particular, the figure shows that predictions of monsoon rainfall based on dynamical models have significant skill in some models. The figure also shows that the skill of statistical monsoon predictions based on the observed May NINO3 index are insignificant.

The superior skill of dynamical models is attributed to (1) the hypothesis that slowly evolving sea surface temperatures are the primary source of monsoon rainfall predictability (Charney and Shukla, 1981), and (2) coupled atmosphere-ocean models produce skillful predictions of sea surface temperatures (Barnston et al., 2012). What is perhaps surprising is that dynamical models make *better* predictions of June-September sea surface temperatures based on antecedent information than statistical models. This fact will be re-confirmed below. This explanation suggest that skillful predictions of monsoon rainfall can be constructed using sea surface temperatures *predicted* by dynamical models. In this paper, we take readily available seasonal predictions of sea surface temperatures and apply regression techniques to make real-time skillful predictions of Indian summer monsoon rainfall. The prediction makes use only of information that is available by early May.

2 Data

The predicted sea surface temperatures (SSTs) used to make predictions of monsoon rainfall come from the NCEP Climate Forecast System v2 (CFSv2) (Saha et al., 2010). Hindcasts for the period 1982-2010 were generated every 5 days at 0, 6,12, 18Z. Real-time forecasts are generated every day, four times a day, for the period 2011-2013. To ensure consistency between the training and real-time forecast data sets, we seek forecasts and hindcasts that are initialized on the same calendar days in early May. The only initial start dates in early May that are common to both the hindcast and forecast data sets are

- 1 May at 0, 6, 12, 18Z
- 6 May at 0, 6, 12, 18Z

Accordingly, we use CFSv2 hindcasts and forecasts initialized only on the above dates. The resulting hindcasts and forecasts are interpreted as an eight-member ensemble. We analyze only the ensemble mean prediction, which is the single best estimate of the predicted state in a mean square sense.

The CFSv2 forecasts were obtained from the National Multi-Model Ensemble project, downloaded from

http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/

The observational SST data used to compare with hindcasts and forecasts is the NOAA Optimal Interpolation Sea Surface Temperature analysis of Reynolds et al. (2007). The NINO3.4 index was chosen for measuring ENSO variability, although no significant difference was found for other standard ENSO indices. The NINO3.4 index is defined as the area average SST in the region bounded by $170°W - 120°W$ and $5°S - 5°N$.

The JJAS Indian Summer Monsoon Rainfall (ISMR) index is defined as the area weighted mean rainfall over India, based on the observational data set produced by the Indian Institute of Tropical Meteorology (IITM) downloaded

ftp://www.tropmet.res.in/pub/data/rain/iitm-subdivrf.txt

The 1871-2010 mean ISMR is 848mm. It should be recognized that this mean differs from that derived from the Indian Meteorological Department (IMD) data set. For comparison, the 1951-2000 mean JJAS all-India area weighted monsoon rainfall is 890cm for the IMD data set, and 845cm for the IITM data set, giving a difference of 45mm, or about 5% of the IMD mean. Thus, the absolute value of the rainfall prediction can be misleading if one does not keep in mind differences between rainfall data sets. To minimize confusion, we report forecasts self-consistently as a percent departure from the long term average of the rainfall data used in this work (i.e., the IITM data set).

It has recently been discovered that the bias of CFSv2 hindcasts has an apparent discontinuity across 1999 (Kumar et al., 2012). This discontinuity has been attributed to the abrupt introduction in October 1998 of Advanced Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (ATOVS) radiance data into the assimilation system (Kumar et al., 2012). Although ATOVS data were assimilated only in the atmospheric model, it influenced the surface forcing fields used to construct ocean initial conditions. Owing to this discontinuity, the hindcasts anomalies for May initial conditions from 1982-1998 were computed by subtracting the 1982-1998 mean, while the anomalies from 1999-2010 were computed by subtracting the 1999-2010 mean. Also, we construct a regression model that accounts for a change in hindcast climatology across 1999, as discussed below.

3 Important Relations from the Hindcast Data Set

To establish a basis for our proposed prediction method, we first confirm that monsoon rainfall is related to ENSO. The correlation between observed ISMR and observed monthly mean NINO3.4 is shown as the black curve in fig. 2. In this figure, ISMR is always evaluated for fixed JJAS season, whereas NINO3.4 is evaluated for the month indicated on the abscissa.

The figure shows that the correlation between observed ISMR and NINO3.4 is weak and insignificant for May and June, but becomes significant for July and thereafter. Two important conclusions can be drawn from this result. First, predictions of monsoon rainfall based on *antecedent* NINO3.4 (i.e., prior to July) have no significant skill. Second, the *simultaneous* correlation between monsoon rainfall and NINO3.4 is significant, suggesting a relation between the two. Unfortunately, this relation is of little predictive value since the NINO3.4 index during the summer monsoon is not available for prediction purposes. However, if the NINO3.4 index during the monsoon season could be predicted accurately in advance, then it is plausible that those predictions of NINO3.4 could be used to predict Indian monsoon rainfall.

The correlation between the observed NINO3.4 index, and the NINO3.4 index predicted by CFSv2 based on early May initial conditions, is shown as the blue curve in fig. 2. In computing this correlation, the hindcasts anomalies for May initial conditions from 1982-1998 were computed by subtracting the 1982-1998 mean, while the anomalies from 1999-2010 were computed by subtracting the 1999-2010 mean. The correlation skill generally decays from a maximum value of 0.91 at the initial month in May, to 0.78 by December, and is highly significant throughout the year. This result demonstrates that the CFSv2 gives skillful predictions of NINO3.4 6-8 months in advance. In contrast, the autocorrelation of the observed NINO3.4 index, with fixed May initial condition, is shown as the green curve in fig. 2. This autocorrelation function is equivalent to the correlation skill of a linear, univariate regression model. The correlation skill of the regression model drops rapidly to 0.5 with lead time. Thus, both the dynamical and linear regression models have statistically significant skill, but at long lead times the skill of the dynamical model is much greater than the regression model skill.

Finally, the correlation between CFSv2-predicted NINO3.4 and observed ISMR is shown as the red curve in fig. 2. These correlations track the corresponding correlations based on observed NINO3.4 (i.e., the black curve), but are systematically weaker, consistent with the loss of skill due to replacing the observed NINO3.4 with predicted NINO3.4. Nevertheless, this weaker correlation is more than compensated by the fact that the predicted NINO3.4 index is available in May, before the summer monsoon season.

4 CFSv2 Forecasts of NINO3.4

The CFSv2 based on May 1 and May 6, 2013 initial conditions predicts a JJAS mean NINO3.4 anomaly of about $0.27^{\circ}K$. The anomaly is computed by subtracting the 1999-2010 mean hindcasts initialized on these start dates from the May 2013 real-time forecasts. Note that the anomaly reported here, based on the eight-member ensemble generated from May 1 and 6 initial conditions, does not correspond to the anomaly predicted in official NCEP forecasts, based on the 28-member ensemble generated from May 1-7 initial conditions.

To put this prediction in context, we show the most recent hindcasts and forecasts of

NINO3.4 in fig. 3. The figure reveals that the NINO3.4 forecasts for the last 3 years were positive. Based on the negative ENSO-monsoon relation indicated in fig. 2, above average NINO3.4 implies that the monsoon is more likely than not going to be below average. More quantitative predictions of monsoon rainfall can be derived from regression methods, as discussed next.

5 The Regression Model

We now develop a regression model for predicting ISMR based on predicted NINO3.4. The first step is to select an appropriate predictor. Fig. 2 shows that the relation between ISMR and predicted NINO3.4 is nearly independent of month, after June. Since there is little basis for choosing a particular month with which to make forecasts for the whole monsoon season, we simply average the NINO3.4 index during the monsoon season, namely between June and September inclusive. Our forecast is based on the assumption that the observed ISMR is related to the *CFSv2-predicted* JJAS NINO3.4 index through a model of the form

$$\text{OBS\_ISMR} = \beta_0\ \text{CFSv2\_NINO3.4} + \beta_1 H(1999 - t) + \beta_2 H(t - 1999) + \epsilon, \qquad (1)$$

where $H(s)$ is a step function defined as

$$H(s) = \begin{cases} 1 & \text{if } s \geq 0 \\ 0 & \text{if } s < 0 \end{cases}, \qquad (2)$$

t denotes the year, β_0, β_1, and β_2 are regression parameters to be estimated from data, and ϵ is a noise term assumed to be independently and identically distributed as a normal distribution with zero mean and variance σ_ϵ^2. The parameters β_1 and β_2 account for possible changes in climatology across 1999. Using the 1982-2010 hindcast data, the parameters β_0, β_1, and β_2 were estimated by the method of least squares, and the parameter σ_ϵ was estimated from the residuals of the model. Details of this calculation are discussed in the appendix. The resulting values are listed in table 1. We also show the regression parameters estimated from the observed simultaneous (i.e. JJAS) NINO3.4 for the model

$$\text{OBS\_ISMR} = \beta_0\ \text{OBS\_NINO3.4} + \beta_1 + \epsilon. \qquad (3)$$

Comparison between columns shows that the slope parameter β_1 drops by about 40% (i.e., from -44mm/K to -26mm/K) when predicted NINO3.4 is substituted for observed NINO3.4. The fact that the parameters in the two columns are comparable to each other suggests that the regression model is detecting the same ENSO-monsoon relation, and that the same regression model could be used for other models with similar skill.

For a $0.27^\circ K$ NINO3.4 anomaly, the above model gives a predicted *mean* ISMR of

$$810mm - 0.27\ *\ 26mm/K \approx 803mm. \qquad (4)$$

parameters	NINO3.4 predicted from CFSv2	NINO3.4 from JJAS OISST
$\hat{\beta}_0$	-26 mm/K	-44 mm/K
$\hat{\beta}_1$	838 mm	826 mm
$\hat{\beta}_2$	810 mm	
$\hat{\sigma}_\epsilon$	81 mm	74 mm

Table 1: Regression parameters for model (1) estimated from the 1982-2010 CFSv2 hindcasts (May initial conditions), and from the JJAS all-India rainfall calculated from the Indian Institute of Tropical Meteorology data set. The caret ($\hat{}$) indicates a sample estimate.

For reference, the 1871-2010 mean ISMR is 848mm, implying that the predicted mean ISMR is about 5% below the mean *of the IITM data set (which differs form the IMD data set)*. For future reference, we note that the prediction can be expressed as

$$\frac{810\ mm - \text{CFSv2\_NINO3.4} * 26\ mm/K}{848\ mm} \approx 96\% - 3\% * \text{CFSv2\_NINO3.4}; \qquad (5)$$

that is, the regression model predicts a 4% reduction even for vanishing NINO3.4 anomaly, due to the fact that the 1999-2010 mean ISMR is 96% of the long term mean, and an additional 3% reduction in ISMR for every degree of NINO3.4 anomaly predicted by CFSv2.

There is considerable interest in constructing probabilistic predictions. Assuming samples are generated by the model (1), then the first three terms in (1) can be associated with the *conditional mean*:

$$E[\text{OBS\_ISMR}|\text{CFSv2\_NINO3.4}] = \beta_0\,\text{CFSv2\_NINO3.4} + \beta_1 H(1999-t) + \beta_2 H(t-1999). \qquad (6)$$

The conditional mean term (6) is associated with the "explained variance" while the term involving ϵ is associated with the "unexplained variance." The total variance is the sum of the variances of these two terms. Consequently, the variance of the conditional mean (6) must be less than the total variance. Some studies artificially inflate the conditional mean prediction (6) so that it has the same variance as the predictand, but this inflation has no sound justification and destroys the proper probabilistic interpretation. In practice, the parameters in (1) are estimated from data, and hence involve additional uncertainty. All sources of uncertainty can be taken into account through the concept of a *prediction interval*, which is an interval in which future observations will fall with a certain probability. The prediction interval assumes that future samples are taken under the same conditions and from the same model (1). The prediction interval is a standard quantity in statistics (for example Seber and Lee, 2003, sec. 5.3). The calculation of this interval is discussed in the appendix and, for all practical purposes, can be approximated as a Gaussian distribution with mean μ_y given by (4), and a standard deviation σ_y calculated in the appendix:

$$\mu_y \approx 803mm \approx 95\% \quad \text{and} \quad \sigma_y \approx 84mm \approx 10\%. \qquad (7)$$

interval (%)	interval (mm)	climatological pdf (Gaussian)	climatological pdf (Empirical)	prediction pdf
$< 90\%$	< 764	16%	24%	32%
$90 - 96\%$	$764 - 814$	19%	17%	23%
$96 - 104\%$	$814 - 882$	31%	38%	27%
$104 - 110\%$	$882 - 933$	19%	7%	11%
$> 110\%$	> 933	16%	14%	6%

Table 2: Probability that ISMR lies in the indicated intervals as predicted by the regression model (1) trained on data for 1982-2010.

Again, we note that the absolute values can be misleading if they are compared (inappropriately) to other rainfall data sets. To put this probabilistic prediction in context, we show in fig. 4 this prediction distribution along with the best fit Gaussian distribution to the ISMR data from 1871-2010 (a histogram of ISMR also is shown). The anticipated shift toward smaller values relative to the climatological distribution is apparent. The spread of the prediction interval is not significantly smaller than the climatological distribution, because the prediction interval accounts for uncertainty due to estimating parameters in addition to the ENSO-monsoon uncertainty embodied in model (1).

To put the above prediction in further context, we note that the 1999-2010 mean ISMR is 810mm, or 96% of the long term mean, so a prediction of 803mm is only 1% lower than the 1999-2010 mean ISMR. Thus, the mean of our probabilistic prediction (i.e., 5% reduction in ISMR) is not atypical of the most recent period.

The above probabilistic prediction also can be characterized by the probability in which ISMR will lie within certain values. An example is shown in table 2 using break values adopted by the Indian Meteorological Department. The table shows that, according to the regression model, the probability of ISMR falling below 96% of the long term mean is 55%, compared to 41% based on the 1871-2010 historical record and compared to 34% based on the best-fit Gaussian to the 1871-2010 historical record. Note the substantial difference between the empirical and best-fit Gaussian percentiles, suggesting that the Gaussian assumption underestimates the probability of below normal rainfall anomalies, and overestimates the probability of above normal rainfall anomalies.

6 Appendix: Least Squares Method and Prediction Intervals

In this section, we given calculation details of the least squares method and prediction intervals. The material described here is standard and discussed in detail in Seber and Lee (2003). The calculations are facilitated by expressing the model (1) in the matrix form

$$\underset{[N\times 1]}{\mathbf{y}} = \underset{[N\times M]}{\mathbf{X}}\ \underset{[M\times 1]}{\boldsymbol{\beta}} + \underset{[N\times 1]\text{'}}{\boldsymbol{\epsilon}} \tag{8}$$

where the dimension of each vector/matrix appears directly under the variable in question. In particular, the model (1) is of this form under the indentifications

$$\mathbf{y} = \begin{pmatrix} \text{OBS\_ISMR(year 1982)} \\ \text{OBS\_ISMR(year 1983)} \\ \vdots \\ \text{OBS\_ISMR(year 1998)} \\ \text{OBS\_ISMR(year 1999)} \\ \vdots \\ \text{OBS\_ISMR(year 2010)} \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \text{CFSv2\_NINO34(year 1982)} & 1 & 0 \\ \text{CFSv2\_NINO34(year 1983)} & 1 & 0 \\ \vdots & \vdots & \vdots \\ \text{CFSv2\_NINO34(year 1998)} & 1 & 0 \\ \text{CFSv2\_NINO34(year 1999)} & 0 & 1 \\ \vdots & \vdots & \vdots \\ \text{CFSv2\_NINO34(year 2010)} & 0 & 1 \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \tag{9}$$

and $N = 29$ and $M = 3$. A standard result is that the least squares estimates of $\boldsymbol{\beta}$ are (Seber and Lee, 2003, sec. 3.1)

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}, \tag{10}$$

and an unbiased estimate of the variance of $\boldsymbol{\epsilon}$ is (Seber and Lee, 2003, sec. 3.3)

$$\hat{\sigma}_\epsilon^2 = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^T \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)/(N - M). \tag{11}$$

Finally, for a prediction based on the $1\times M$ predictor variable $\mathbf{x}_0$, the $(1-\alpha)100\%$ prediction interval is given by (Seber and Lee, 2003, sec. 5.3)

$$\mathbf{x}_0\hat{\boldsymbol{\beta}} \pm t_{\alpha/2}\hat{\sigma}_\epsilon\sqrt{1 + \mathbf{x}_0\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}_0^T}, \tag{12}$$

where $t_{\alpha/2}$ is the value of t such that area $\alpha/2$ lies to the right of this value for a t distribution with $N - M$ degrees of freedom. For the 1982-2010 period, the sample size is $N = 29$, which is sufficiently large that for all practical purposes the t-distribution can be replaced by the standardized normal distribution. Thus, the probabilistic prediction for our regression model is a normal distribution with mean μ_y and variance σ_y^2 given by

$$\mu_y = \mathbf{x}_0\hat{\boldsymbol{\beta}} \tag{13}$$

$$\sigma_y^2 = \hat{\sigma}_\epsilon^2\left(1 + \mathbf{x}_0\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}_0^T\right). \tag{14}$$

For the 2013 prediction, $\mathbf{x}_0 = \begin{pmatrix} 0.27^\circ K & 0 & 1 \end{pmatrix}$.

References

Barnston, A., M. K. Tippett, M. L. L'Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Am. Meteor. Soc.*, **93**, ES48–ES50.

Charney, J. G. and J. Shukla, 1981: Predictability of monsoons. *Monsoon Dynamics*, J. Lighthill and R. P. Pearce, Eds., Cambridge University Press, 99–109.

DelSole, T. and J. Shukla, 2012: Climate models produce skillful predictions of Indian summer monsoon rainfall. *Geophys. Res. Lett.*, **39**, L09 703, doi:10.1029/2012GL051 279.

Kumar, A., M. Chen, L. Zhang, W. Wang, Y. Xue, C. Wen, L. Marx, and B. Huang, 2012: An analysis of the nonstationarity in the bias of sea surface temperature forecasts for the NCEP Climate Forecast System (CFS) version 2. *Mon. Wea. Rev.*, **140**, 3003–3016.

Rajeevan, M., C. K. Unnikrishnan, and B. Preethi, 2011: Evaluation of the ENSEMBLES multi-model seasonal forecasts of Indian summer monsoon variability. *Clim. Dyn.*, DOI: 10.1007/s00 382–011–1061–x.

Reynolds, R., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *J. Climate*, **20**, 5473–5496.

Saha, S., et al., 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteor. Soc.*, **91**, 1015–1057.

Seber, G. A. F. and A. J. Lee, 2003: *Linear Regression Analysis*. Wiley-Interscience, 557 pp.

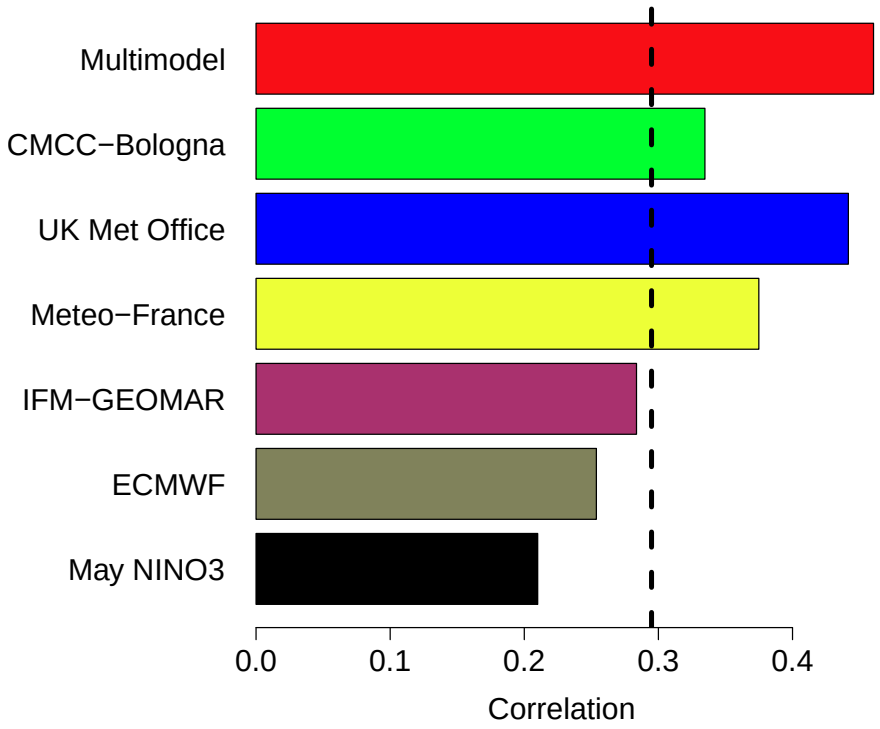

Figure 1: Correlation between observed and predicted JJAS all-India rainfall for hindcasts in the ENSEMBLES data set for the period 1960-2005. All-India rainfall in dynamical models is defined as the total land precipitation within 70E - 90E and 10N - 25N. The 5% significance threshold is indicated by the vertical dashed line. The ensemble mean ISMR was used from each model, while the mean of all model ISMR values was used to compute the "multimodel" value. The correlation between observed rainfall and the May NINO3 index also is shown.

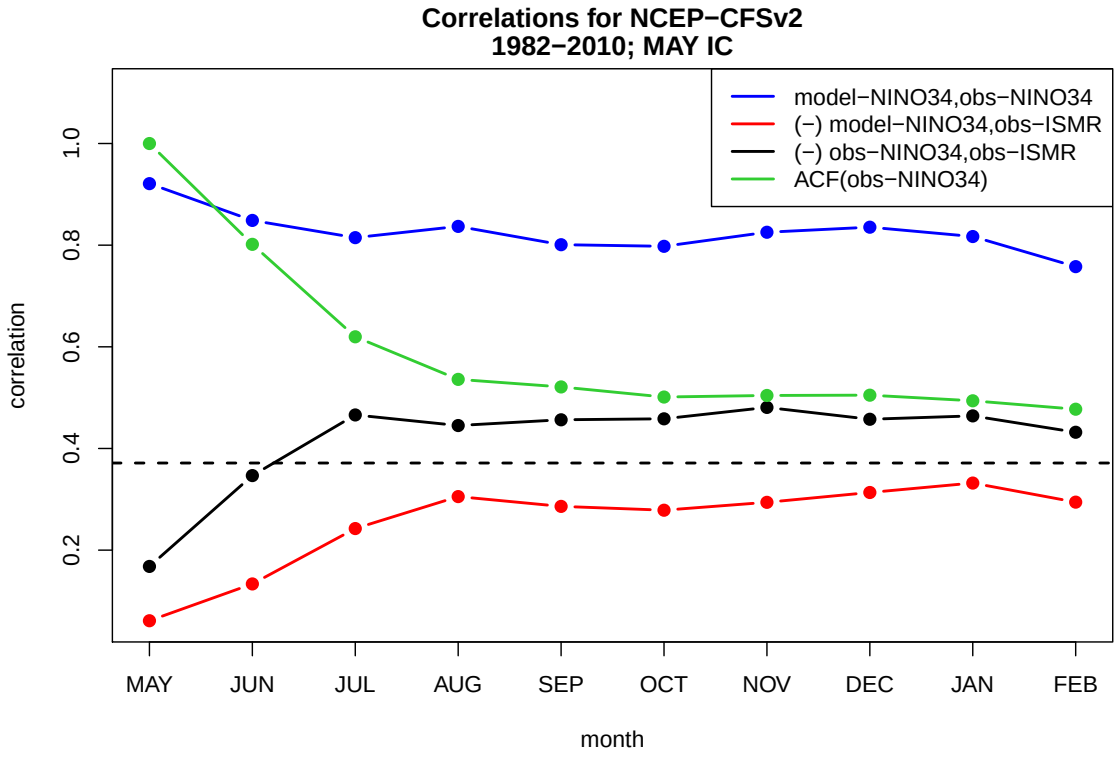

Figure 2: Various correlations between the JJAS Indian Summer Monsoon Rainfall (ISMR) index and NINO3.4 index for 1982-2010. The correlation skill of NINO3.4 hindcasts from CFSv2, based on May initial conditions, is shown in blue. The observed autocorrelation of NINO3.4 using only data starting from May (green). The (negative) correlation between the CFSv2-predicted NINO3.4 and observed ISMR is shown in black. The (negative) correlation between observed ISMR and observed NINO3.4 index is shown in red. The horizontal dashed line shows the 5% significance level. The ISMR index is always evaluated for JJAS, but the NINO3.4 index is evaluated at the month indicated on the abscissa.

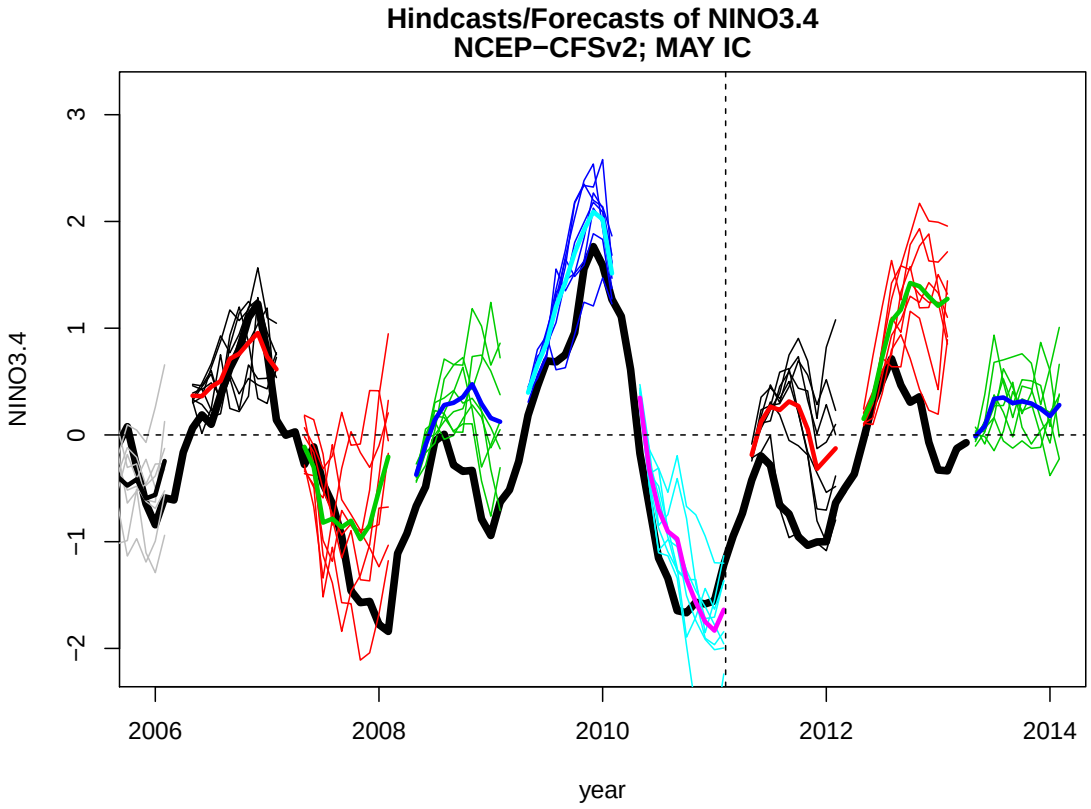

Figure 3: Hindcasts and forecasts of NINO3.4 by CFSv2 for May initial conditions. The observed NINO3.4 anomaly index is shown as the continuous, black curve. Ensemble members for the same initial month have the same color. The ensemble mean is indicated by a thicker curve of different color. The vertical dashed line separates hindcast and forecast years.

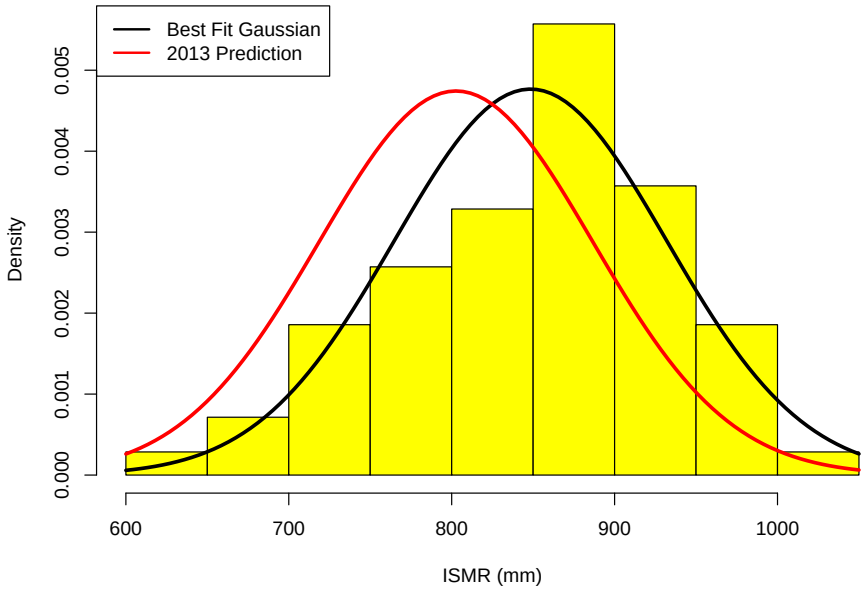

Figure 4: Histogram of 1871-2010 ISMR, the best fit Gaussian to the histogram (black curve), and the 2013 probabilistic prediction based on a regression model estimated from the 1982-2010 period (red).