The City College of New York

Toolbox for Evaluating Ensembles Using an Information Gain Measure

Hannah Aizenman, Michael Grossberg, Irina Gladkova, Nir Krakauer

City College of New York

Abstract

We transformed monthly ensemble temperature predictions from the NCEP Climate Forecast System (CFSv2) into probabilistic prediction using various measures of climatology and forecast uncertainty. By using information gain to evaluate the accuracy of the probabilistic predictions based on these ensembles, we were able to diagnose specific problems in the ensembles. We learned that the standard deviation of past temperatures (climatology) is a better measure of forecast uncertainty than the ensemble spread. These results led us to implement a 2-3 month lookback auto-regressive statistical model to quantify the degree to which climate dynamics in CFSv2 improve forecast skill. We found that combining climatology, recent history, and CFSv2 forecast did better than any individual component. Using the scientific Python ecosystem, we built a library for developing, evaluating, and visualizing these probabilistic forecasts. Our toolbox can be easily adapted to other ensembles.

Quantify Forecast Accuracy

The NCEP CFSv2 predicts weather. Ensembles of CFSv2 monthly hindcasts (Forecasts or f_i) cover about 30 years on an approximately $1^\circ x 1^\circ$ spatial grid.

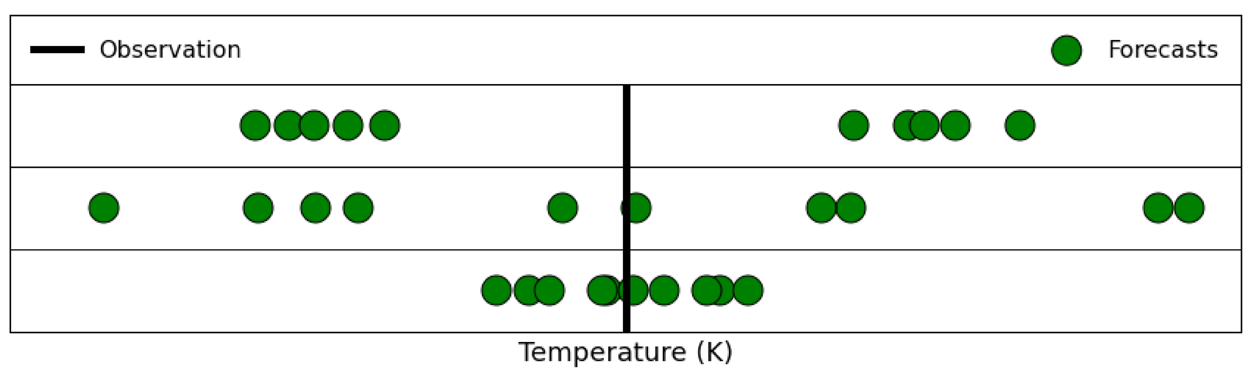

Figure : These ensemble predictions all have a root mean square error (RMSE) of 0.

An observation where the prediction is very spread out (very uncertain) but centered around the mean may have a better RMSE than a much more certain but slightly off center prediction due to RMSE's sensitivity to relative distance from the mean.

	RMSE Relative to Observations (K)
Climatology	1.73
μ(Forecasts)	2.21
Bias Corrected Forecast Mean	1.77

It is hard to asses whether climatology's lower RMSE is due to a more accurate prediction, or if climatology is just well distributed around the observation; therefore we developed a probabilistic method of evaluating skill that incorporates certainty.

Create Probabilistic Models

We estimated the forecast distribution (p^f) using the ensemble members from a given month and year (t), whereas the climatology distribution (p^c) used the climatology from that same month for all prior years; both were computed at a single latitude and longitude. We then computed the probability of the observation occurring in the distribution $p^d(O_{t,lat,lon})$.

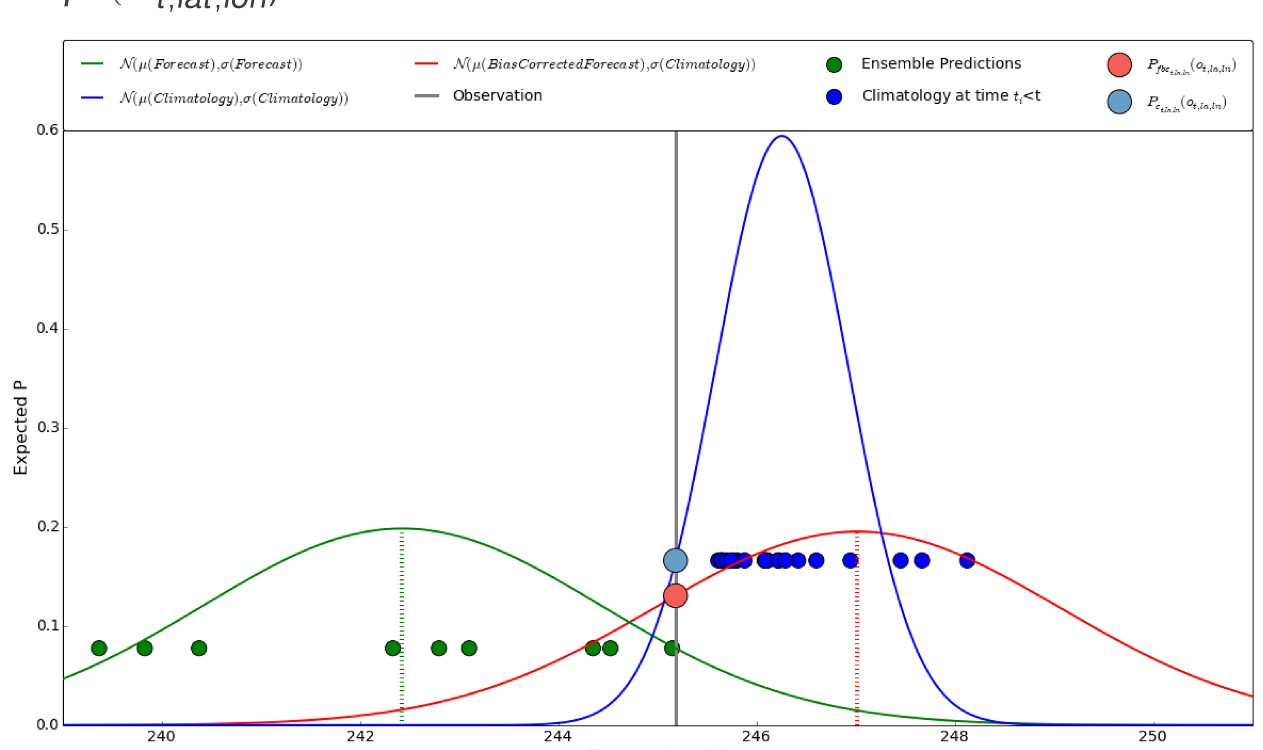

Figure : Estimating p^d of temperature on June, 2003 in the Equatorial Pacific Ocean (-84S, 50E)

We used the probabilistic models $\mathcal{N}(\mu(X), \sigma(Y))$, which are Gaussian distributions with μ given as the mean of the X data and σ as the standard deviation of the Y data.

Acknowledgments

The authors gratefully acknowledge support from NOAA under grants NA11SEC4810004 and NA12OAR4310084. All statements made are the views of the authors and not the opinions of the funding agency or the U.S. government.

Compute Negative Log Likelihood

The observation was interpreted as an informational signal with a bit rate given by the terms of entropy, computed by taking the negative log likelihood (NLL) of $p^d(O_{t,lat,lon})$. To compute the information gain, we first looked at the NLL of the probability of the observation occurring in the model $-\log_2(p^d(O_{t,lat,lon}))$.

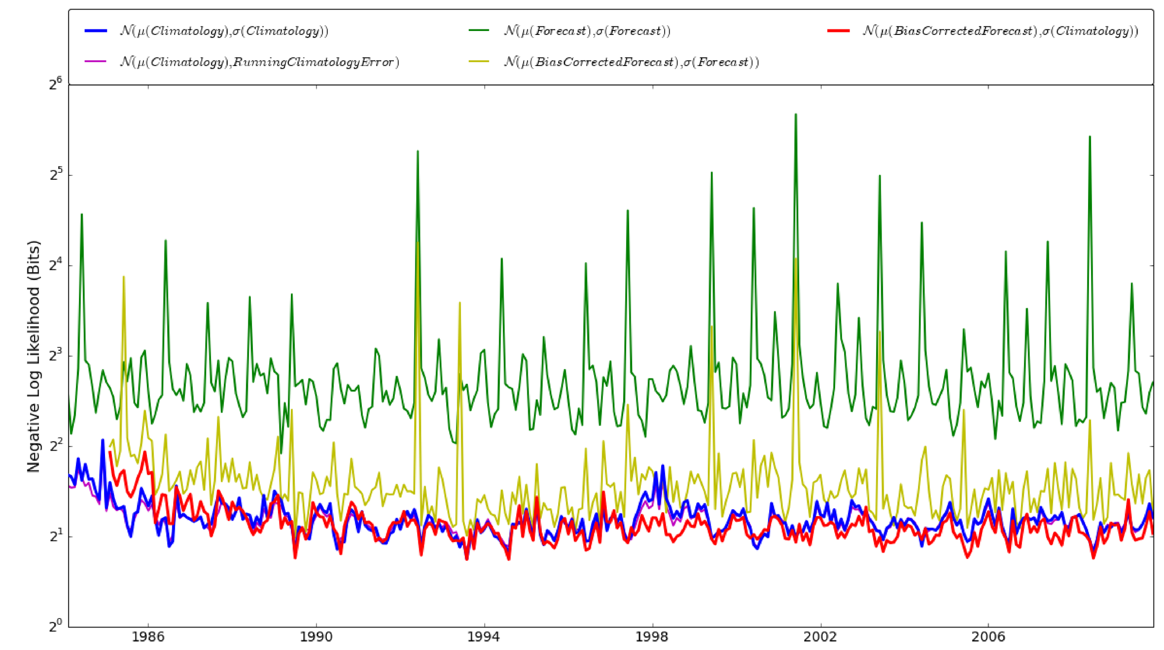

Figure : A good prediction has low NLL, whereas a bad prediction has high NLL.

Measure Information Gain

The information gain is the difference in the NLL of the baseline model and the model being evaluated:

$$IG = -\log_2(P^b(O_{t,lat,lon})) - -\log_2(P^d(O_{t,lat,lon}))$$

IG Relative to $\mathcal{N}(\mu(Climatology), \sigma(Climatology))$	Mean IG	Median IG
$\mathcal{N}(\mu(Climatology), ClimatologyRunningError)$	0.031	-0.107
$\mathcal{N}(\mu(Forecast), \sigma(Forecast))$	-4.898	-0.42
$\mathcal{N}(\mu(BiasCorrectedForecast), \sigma(Forecast))$	-0.88	0.098
$\mathcal{N}(\mu(BiasCorrectedForecast), \sigma(Climatology))$	0.057	0.0152

Table : Larger positive information gain means the model is better at predicting observations

$\mathcal{N}(\mu(Climatology), \sigma(Climatology))$ was chosen as the baseline because it is based only on historical data. The skill of the Bias-Corrected Forecast with Climatology model indicates that the standard deviation of the climatology represents the uncertainty in temperatures across time, latitude and longitude better than the other techniques we used to quantify uncertainty.

Analyze Spatial Distribution of Information Gain

Since we found that the $\mathcal{N}(\mu(BiasCorrectedForecast), \sigma(Climatology))$ model has the highest skill of the ones we tested, we took the mean information gain across all observations at a given latitude and longitude to detect where the forecasts are not doing as well as the baseline (climatology).

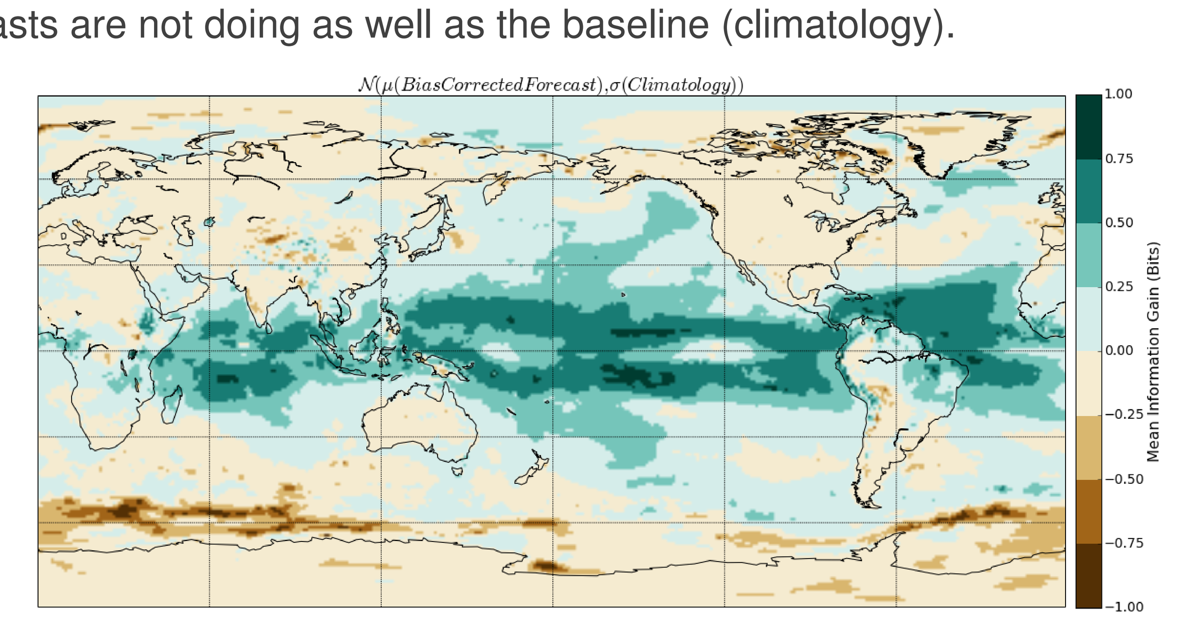

Figure : From this figure, we see that a probabilistic model based on the mean of the ensembles and the standard deviation of the climatology does well in the Equatorial Pacific and over the oceans, but poorly on most landmasses. The model also does poorly in Antarctica and the coast, but that may be due to known problems with obtaining accurate observational data in Antarctica.

Because the IG is spatially inconsistent, we developed a model that could use observational data from a month or two back instead of the smoother climatology signal. Knowing that the CFSv2 forecast and climatology seem to trade off where they do well, we then incorporated forecasts into our linear auto-regressive model. We implemented the model as a simple linear regression where the training set grows to include all observations 3 months prior to the one being predicted.

Combine Climatology and Forecasts in a New Model

The mean of our distribution was obtained using a linear auto-regressive model:

$$\mu_{cr}(t,q) = \alpha(t,q)\mu_c(m(t),q) + \beta_1(t,q)o(t-2,q) + \beta_2(t,q)o(t-3,q)$$

where $\alpha(t,q)$, $\beta_1(t,q)$, and $\beta_2(t,q)$ were fit for all times $t' < t-1$ at location q. We represented the uncertainty (which was used as the standard deviation) as the running error:

$$\mathbf{err}_{cr}(t',q) = \mu_{cr}(t',q) - o(t',q)$$

Forecasts were added to the regression as the γ coefficient:

$$\mu_{fr}(t,q) = \alpha(t,q)\mu_c(m(t),q) + \beta_1(t,q)o(t-2,q) + \beta_2(t,q)o(t-3,q) + \gamma\mu_f(t,q)$$

$$\sigma_{fr}(t,q) = \mathbf{stdev}(\{\mathbf{err}_{fr}(t',q)\}_{t'<t}).$$

Evaluate Information Gain of Regression Based Models

We accounted for the effect of trends in the data by computing climatology as an exponentially weighted moving average (EWMA), weighing the observations in the training data using EWMA, and with a combination of both these methods.

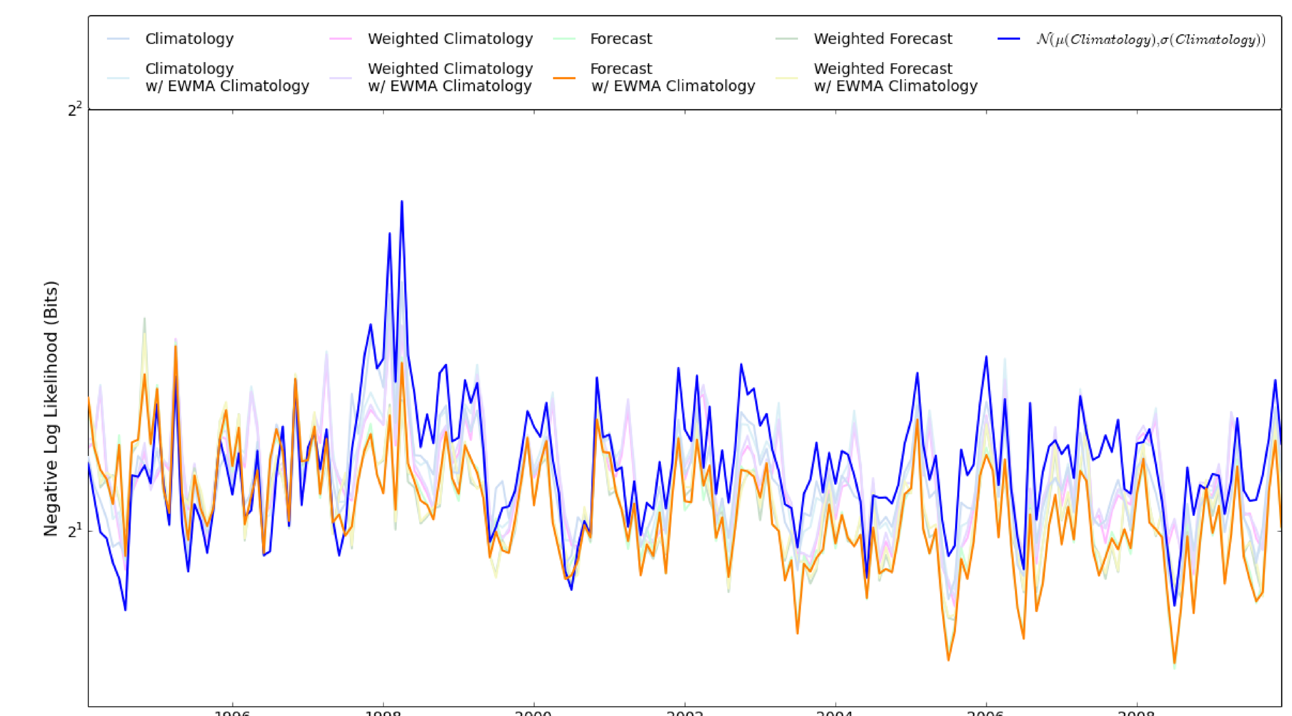

Figure : The data was trimmed 10 years to let the model stabilize.

IG Relative to $\mathcal{N}(\mu(Climatology), \sigma(Climatology))$	Mean IG	Median IG
Climatology Regression	0.064	-0.058
Climatology: EWMA Climatology	0.032	-0.144
Climatology: Weighted	0.056	-0.126
Climatology: Weighted and EWMA Climatology	0.039	-0.155
Forecast Regressions	0.158	0.014
Forecast: EWMA Climatology	0.16	0.009
Forecast: Weighted	0.1421	-0.011
Forecast: Weighted and EWMA Climatology	0.134	-0.029

Table : Larger positive information gain means the model is better at predicting observations

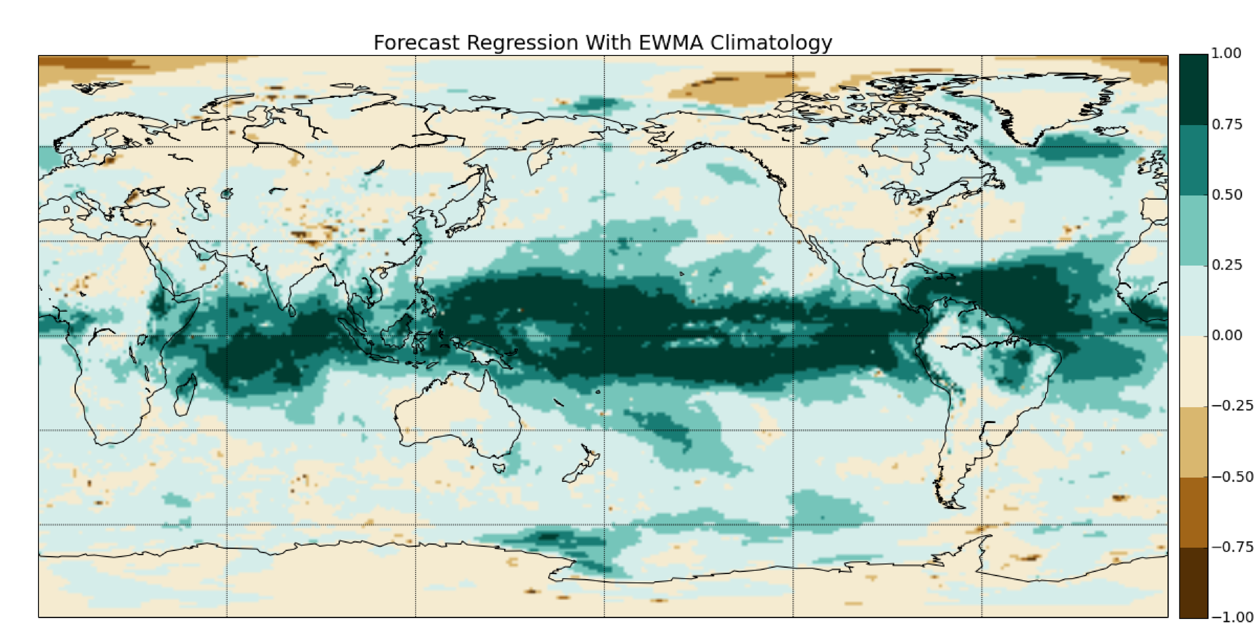

Figure : Prediction skill on land increased when forecast and EWMA climatology were added into the model.

Conclusion

Through use of our toolbox, we learned that an accurate representation of uncertainty improves probabilistic models and that model skill varies spatially. Using this knowledge, we developed a new model that incorporates uncertainty, climatology, past observations, and forecasts. We then evaluated this model using our toolkit and found gains in predictions over land, especially in the equatorial region. Since the only assumption our toolkit makes is that the data is spatio-temporal, it can be readily used to evaluate skill in predicting other physical variables and adapted to work with other ensemble systems.