

# Statistics 220

## Bayesian Data Analysis

Mark E. Irwin  
Department of Statistics  
Harvard University

Spring Term

---

Thursday, February 3, 2005 - Tuesday, May 17, 2005



# Personnel

Instructor: Mark Irwin  
Office: 611 Science Center  
Phone: 617-495-5617  
E-mail: [irwin@stat.harvard.edu](mailto:irwin@stat.harvard.edu)

Lectures: Tues, Thurs 10:00 - 11:30, Science Center 221  
Office Hours: Monday & Wednesday, 11:00 - 12:00  
or by appointment

Teaching Fellow: Alan Lenarcic  
E-mail: [lenarcic@fas.harvard.edu](mailto:lenarcic@fas.harvard.edu)  
Office Hours: To be announced

## Text Book

Required Text: Gelman A, Carlin JB, Stern HS, & Rubin DB (2004)  
Bayesian Data Analysis, 2nd edition.  
Chapman & Hall/CRC.  
Available from the COOP and on reserve in Cabot Library

Suggested  
References: Gilks WR, Richardson S, and Spiegelhalter DJ (1996).  
Markov Chain Monte Carlo in Practice.  
Chapman & Hall / CRC.

Liu J (2001). Monte Carlo Strategies in Scientific  
Computing. Springer-Verlag.

# Computing

Many of the computational problems can be done in BUGS (Bayesian inference Using Gibbs Sampling). Windows and Linux (Intel based) versions are available. Familiarity with BUGS is not assumed. R/S-Plus will also be useful.

Links to software and other computing resources are available from the course web site or at

<http://www.people.fas.harvard.edu/~mirwin/computing/>

# Grading

Course grades will be determined based on homework assignments (4 or 5 during the term) and a term project, the exact form of which is to be determined.

# Web Site

<http://www.courses.fas.harvard.edu/~stat220/>

Available on the site:

- PDF files of the lecture notes
- Data files used in class, in text format
- References, with many of the papers available in pdf format.
- Assignments, solutions, data files for them
- Class announcements, calendar, etc

# Course Goals

## 1. Basics of the Bayesian inference:

- (a) setting up a probability model and the basics of statistical inferences;
- (b) using the probability theory and the Bayes rule;
- (c) simple examples: binomial, normal, and exponential families;
- (d) non-informative priors;
- (e) exchangeability and modelling;
- (f) asymptotic normality (informal introduction).

## 2. Multi-parameter models, normal with unknown mean and variance, the multivariate normal distribution, multinomial models, Dirichlet process models, and examples. Computation and simulation from arbitrary posterior distributions in two parameters.

3. Inference from large samples and comparison to standard non-Bayesian methods.
4. Hierarchical models, estimating several means, estimating population parameters from data, meta analysis, and some examples. Emphasis on statistical thinking for complicated problems.
5. Model checking, comparison to data and prior knowledge, sensitivity analysis, prior and posterior predictive checking, posterior odds ratio, model likelihood.
6. Computation 1 (non-iterative methods): integration, conjugate analysis, normal and Laplace approximations, importance sampling, rejection sampling, sequential imputation.
7. Computation 2 (iterative sampling methods): Metropolis-Hastings algorithms and the Gibbs sampler. Their variations include Metropolized independent sampling, iterative predictive updating, and simulated tempering; etc.

8. Missing data problems and data augmentation methodology. Latent variable models; Student-t models for robust inference and sensitivity analysis; Bayesian nonparametrics.

## Who is this man?

This is reportedly the only known picture of the Reverend Thomas Bayes, F.R.S. — 1701? - 1761.

This picture was taken from the 1936 History of Life Insurance (by Terence O'Donnell, American Conservation Co., Chicago). As no source is given, the authenticity of this portrait is open to question.

So what is the probability that this is actually Reverend Thomas Bayes?

However there is some additional information. How does this change our belief about who this is?



Many details about Bayes are sketchy. According to Steve Stigler

The date of his birth is not known: Bayes's posterior is better known than his prior.



Actually his date of death isn't that well known either. It's generally considered to be April 7, 1761 however it has also been reported as April 17th of the same year.

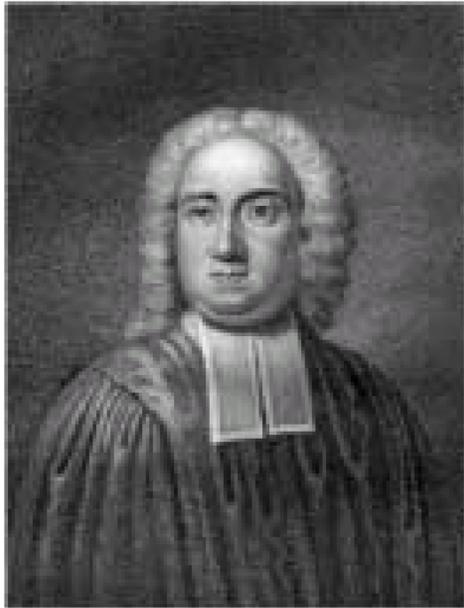
There is some additional information we can get from this picture to help us decide whether this is Bayes or not.

1. The caption under the photo in O'Donnell's book was "Rev. T. Bayes: Improver of the Columar Method developed by Barrett".

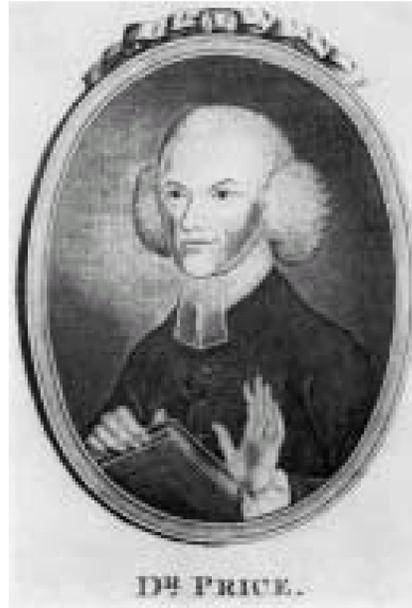
There are some problems with this claim. First Barrett was born in 1752 and would have been about 9 years old when Bayes died. In addition, the method that Bayes allegedly improved was apparently developed between 1788 and 1811 and read to the Royal Society in 1812, long after Bayes' death.

So there is a problem here, but whether it is a problem with just the caption or the picture as well isn't clear.

2. Bayes was a Nonconformist (Presbyterian) Minister. Does the clothing in the picture match that of a Nonconformist Minister in the 1740's and 1750's. The picture has been compared to three other Ministers, Joshua Bayes, Bayes' father, Richard Price (portrait dated 1776), the person who read Bayes' paper to the Royal Society, and Philip Doddridge, a friend of Bayes' brother-in-law.



Joshua Bayes  
(1671-1746)



Richard Price  
(1723-1791)



Philip Doddridge  
(1702-1751)

Two things stand out in the comparisons.

- (a) No wig. It is likely that Bayes should have been wearing a wig similar to Doddridge's, which was going out of fashion in the 1740's or similar to Price's, which was coming into style at the time.
- (b) Bayes appears to be wearing a clerical gown like his father or a larger frock coat with a high collar. On viewing the other two pictures, we can see that the gown is not in the style for Bayes' generation and the frock coat with a large collar is definitely anachronistic.

(Interpretation of David Bellhouse from IMS Bulletin **17**, No. 1, page 49)



Question: How to we incorporate this information to adjust our probability that this is actually a picture of Bayes?

Answer:  $P[\text{This is Bayes}|\text{Data}]$  which can be determined by Bayes' Theorem.

$$P[B|\text{Data}] = \frac{P[B]P[\text{Data}|B]}{P[B]P[\text{Data}|B] + P[B^c]P[\text{Data}|B^c]}$$

Note that this is not easy to do as assigning probabilities here is difficult.

For an example of how this can be done, see the paper (available on the course web site)

Stigler SM (1983). Who Discovered Bayes's Theorem. *American Statistician* **37**: 290-296.

In this paper, Stigler examines whether Bayes was the first person to discover what is now known as Bayes' Theorem. There is evidence that the result was known in 1749, 12 years before Bayes' death and 15 years before Bayes' paper was published. In Stigler's analysis

$$P[\text{Bayes 1st discovered} \mid \text{Data}] = 0.25$$

$$P[\text{Saunderson 1st discovered} \mid \text{Data}] = 0.75$$

Note that Bayes actually proved a special case of Bayes' Theorem involving inference on a Bernoulli success probability.

An example where Bayes' approach could be used is . . .

# How accurate are Punxsutawney Phil and Wiarton Willie?



Punxsutawney Phil  
Shadow in 2005



Wiarton Willie  
No Shadow in 2005

So what are the probabilities that Punxsutawney Phil and Wiarton Willie make accurate predictions?

According to the Wiarthon Willie website, Willie has been correct 37 times out of 41 years ( $\hat{p} = \frac{37}{41} = 0.902$ ).

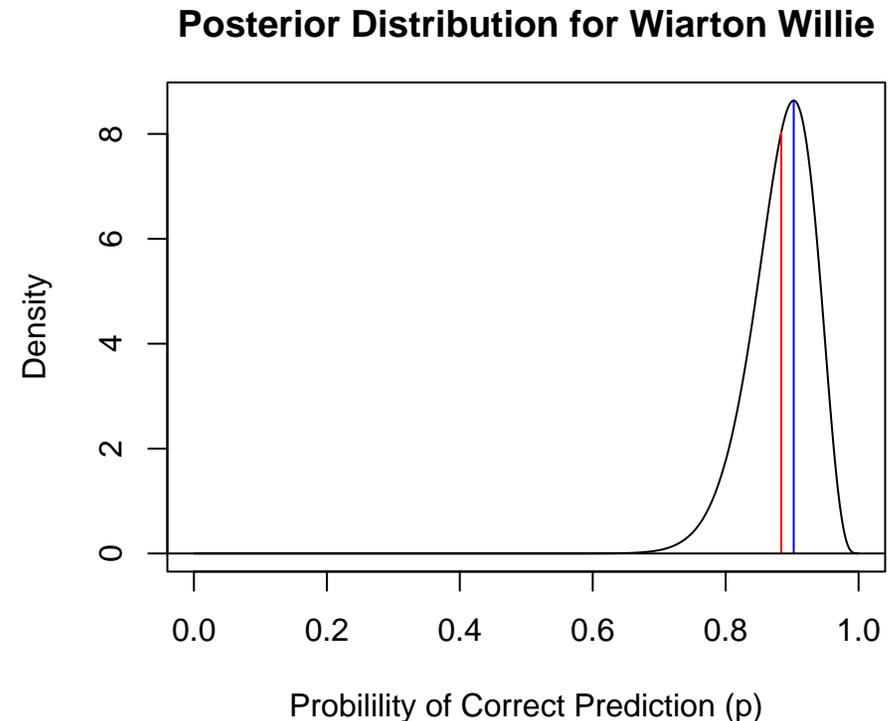
By Bayes' approach

$$E[p|\text{Data}] = \frac{38}{43} = 0.884$$

$$SD(p|\text{Data}) = 0.048$$

$$\text{Mode}(p|\text{Data}) = \frac{37}{41} = 0.902$$

(I suspect their reported results are a bit biased. :) The only numbers I could find for Punxsutawney Phil say he has never been wrong.)



# Statistical Inference

Two main approaches

- Frequentist (Stat 111/211)

Model parameters are fixed unknown quantities. Randomness only in data.

- Estimation - Maximum likelihood, method of moments
- Confidence intervals
- Significance testing -  $p$ -values
- Hypothesis testing - Reject/Don't Reject  $H_0$

- Bayesian (Stat 220)

Model parameters are random variables. Inference is based on  $P[\theta|\text{Data}]$ , the posterior distribution given the data.

- Estimation - Posterior means, modes
- Credible intervals/sets
- Posterior probabilities

There is also decision theory which tends to be Bayesian in nature (good procedures (e.g Admissible procedures) must be Bayesian with respect to some prior).

# Why Should You Use Bayesian Procedures?

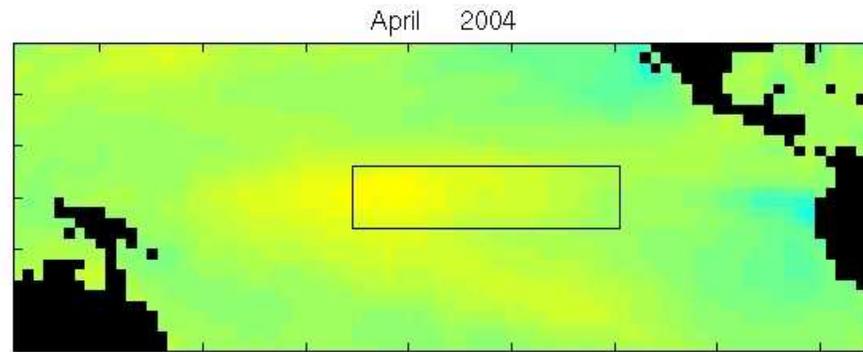
- It is the only the correct form of inference. :)

Other inference procedures can be suboptimal or in fact give inconsistent answers.

- Allows for a consistent method to combine prior beliefs and data for inference procedures.
- Usually have good frequentist properties. A 95% credibility interval will often have coverage level close to 95%.
- Can give answers in large, high dimensional problems when frequentist solutions will tend to break down.

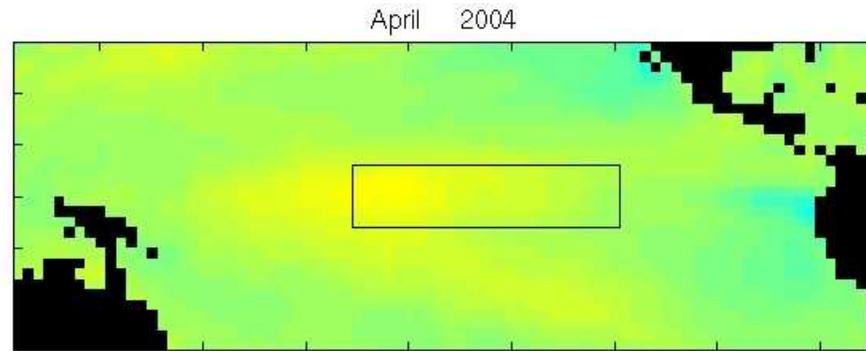
## Example: Tropical Pacific Sea Surface Temperature Evolution

Probabilistic based model involving wind, sea surface temperature, and the Southern Oscillation Index (air pressure measure) data to forecast sea surface temperatures 7 months in the future.



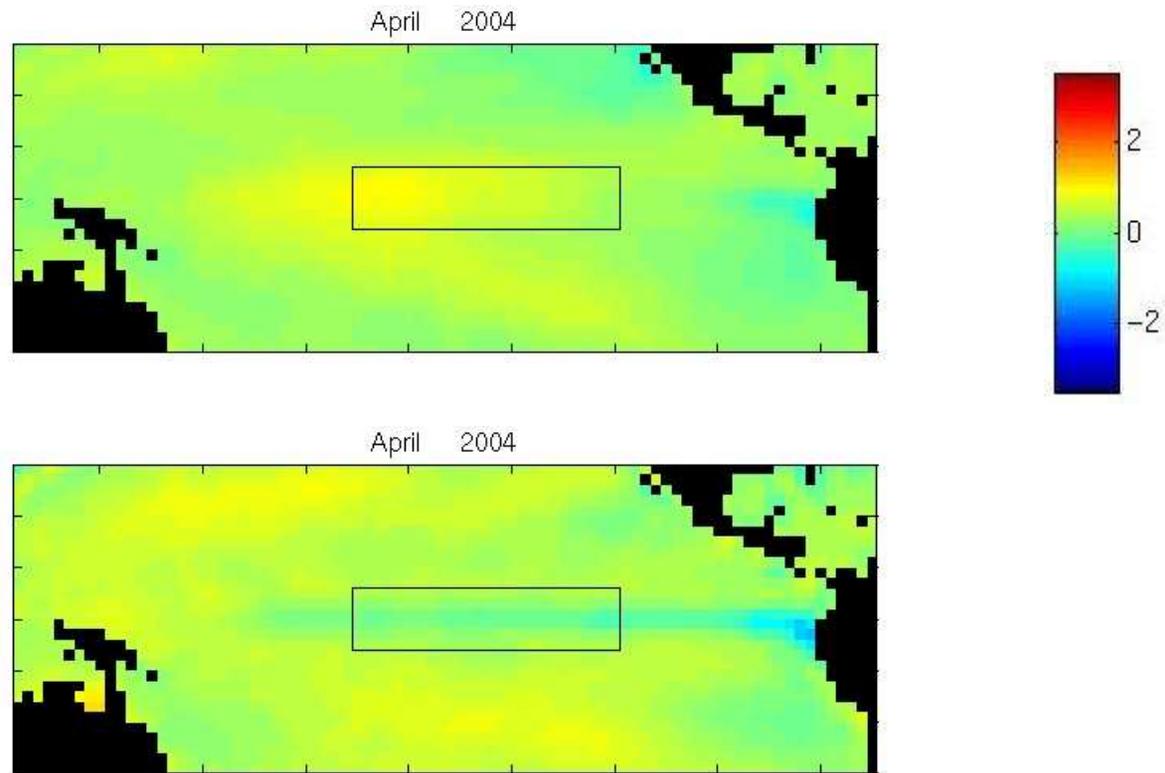
A highly complex, hierarchical model was used. In addition to the three sets of input variables into the model, there is a regime component, recognizing that the region goes through periods of below normal, above normal, and typical temperatures. As part of the analysis, we can ask what would temperature field look like conditional on being in a given temperature regime.

Currently there has been 540 months (January 1970 to December 2004) of data analyzed with temperatures forecast on a  $30 \times 84$  (= 2520) pixel grid.



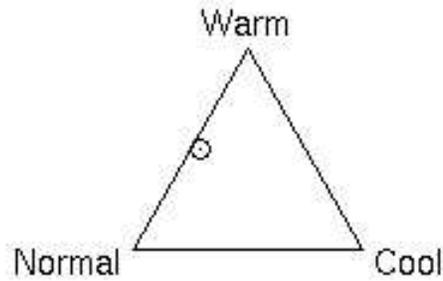
The forecast map is a summary of the probabilistic model (the posterior mean), which is a weighted average over all possible sea surface temperature maps. As we can determine the likelihood of any temperature field, summary measures of the uncertainty in the forecasts can be determined.

The calculation of the posterior distributions was done by Markov Chain Monte Carlo (MCMC) based on 1000 imputations with a burn in of 100 imputations (I think).

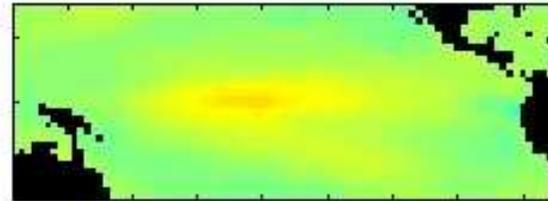


April 2004 anomaly forecast (top) and observed anomaly (bottom) based on Jan 1970 to September 2003 data.

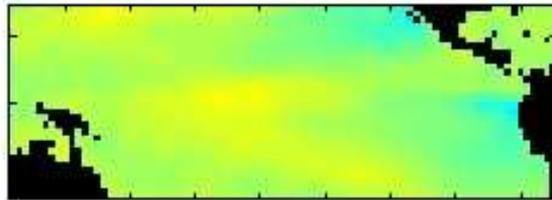
April 2004



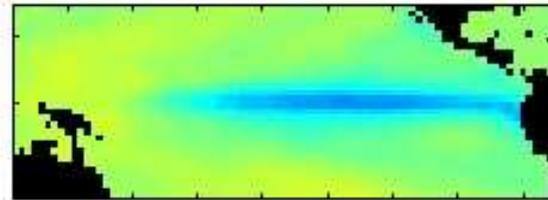
Warm: Prob = 0.501



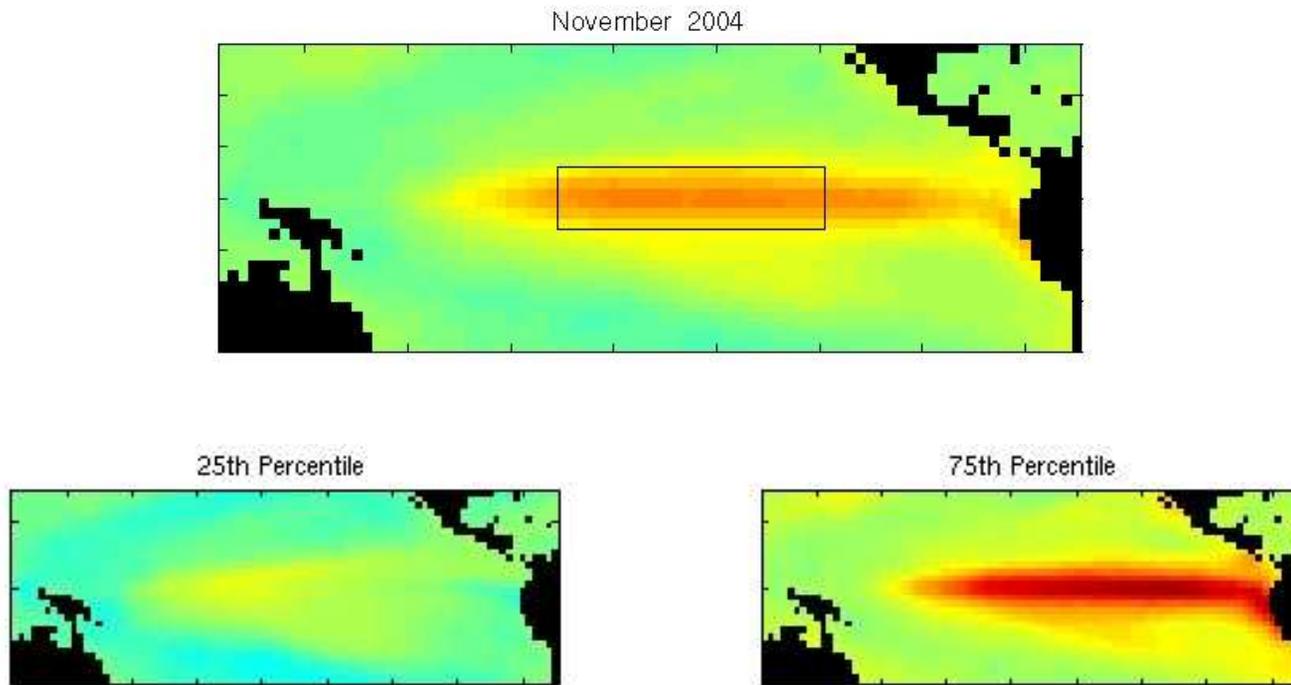
Normal: Prob = 0.460



Cool: Prob = 0.038

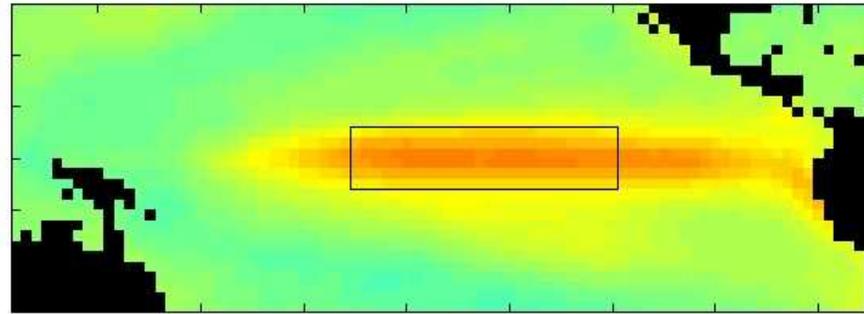


April 2004 regime specific anomaly forecasts based on Jan 1970 to September 2003 data.



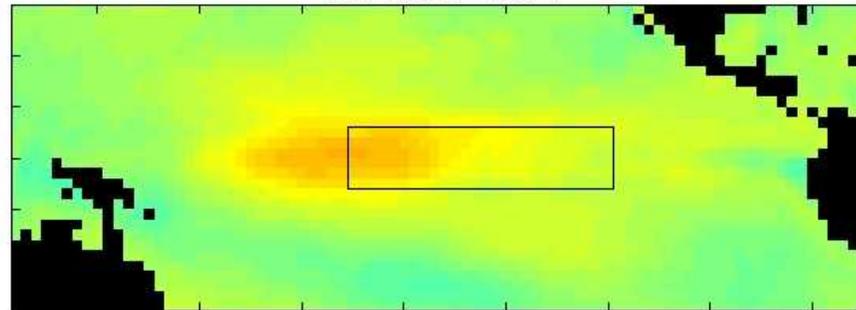
November 2004 anomaly forecast with 25th and 75th percentiles based on Jan 1970 to April 2004 data.

November 2004



Niño 3.4 Forecast =  $1.375^{\circ}\text{C}$

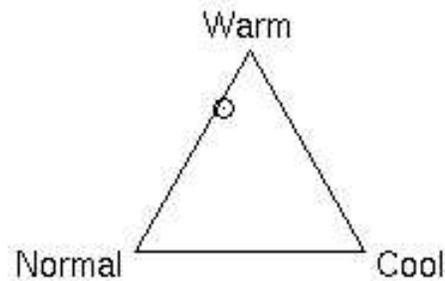
November 2004



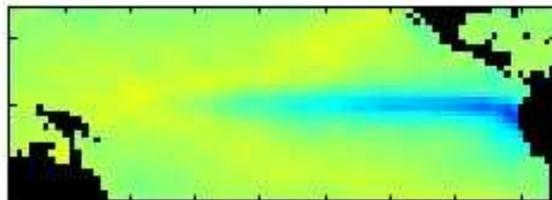
Niño 3.4 Observed =  $0.851^{\circ}\text{C}$

November 2004 anomaly forecast (top) and observed anomaly (bottom) based on Jan 1970 to April 2004 data. This forecast suggests an El Niño this winter.

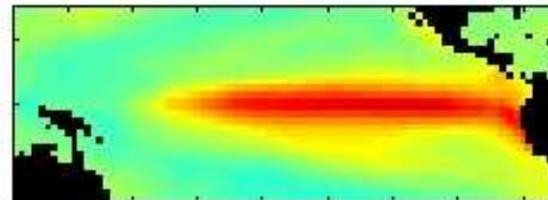
November 2004



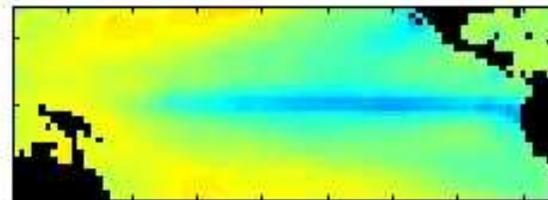
Normal: Prob = 0.253



Warm: Prob = 0.721



Cool: Prob = 0.027



November 2004 regime specific anomaly forecasts based on Jan 1970 to April 2004 data.

Refs:

Berliner, Wikle, and Cressie (2000). Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling. *Journal of Climate*, **13**, 3953-3968.

[http://www.stat.ohio-state.edu/~sses/collab\\_enso.php](http://www.stat.ohio-state.edu/~sses/collab_enso.php)

Note that the model used here is only an approximation of reality. This does not include information about the underlying physics of the meteorology.